

Universidade do Minho
Escola de Ciências

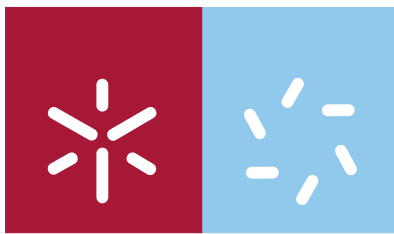
Liliana da Silva Matos

**Análise funcional de uma mutação *missense*
no gene *IDS* associada a alterações
de *splicing*.**

Liliana da Silva Matos

UMinho | 2009

Dezembro de 2009



Universidade do Minho
Escola de Ciências

Liliana da Silva Matos

**Análise funcional de uma mutação *missense*
no gene *IDS* associada a alterações
de *splicing*.**

Tese de Mestrado em Genética Molecular

Trabalho efectuado sob a orientação da
Doutora Sandra Alves
e do
Professor Doutor Rui Oliveira

Dezembro de 2009

DECLARAÇÃO

Nome: Liliana da Silva Matos

Endereço electrónico: lilma81@gmail.com **Telefone:** 91 3029317

Número do Bilhete de Identidade: 12004283

Título dissertação: Análise funcional de uma mutação *missense* no gene *IDS* associada a alterações de *splicing*.

Orientador(es): Doutora Sandra Alves e Professor Doutor Rui Oliveira

Ano de conclusão: 2009

Designação do Mestrado: Mestrado em Genética Molecular

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE/TRABALHO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE;

Universidade do Minho, 29 de Dezembro de 2009

Assinatura:

Dissertação apresentada à Universidade do Minho para cumprimento dos requisitos necessários à obtenção do grau de mestre em Genética Molecular, realizada na Unidade de Investigação e Desenvolvimento do Centro de Genética Médica Doutor Jacinto de Magalhães – INSA, I.P, Porto, sob a orientação da Doutora Sandra Alves, Investigadora Auxiliar do CGM-INSA, e do Professor Doutor Rui Oliveira, Professor Auxiliar da Universidade do Minho.

*Aos meus pais, que me deram a vida,
os seus genes e tudo aquilo que
apenas meras palavras não permitem
descrever*

Agradecimentos

O presente trabalho contou com o apoio e colaboração de várias pessoas que, de uma forma ou de outra, contribuíram para a sua realização. Gostaria, portanto, de expressar a todos o meu muito obrigada!

Contudo, gostaria de agradecer de uma forma especial:

À Doutora Sandra Alves por me ter dado a oportunidade de integrar o seu grupo de investigação, pela orientação deste trabalho, pela confiança que depositou nas minhas capacidades e pela autonomia que me deu, tendo permitido efectuar todo o trabalho ao meu ritmo. Agradeço ainda todos os conhecimentos transmitidos e todo o rigor científico e sugestões que ajudaram a melhorar a redacção desta tese. Obrigada por toda a disponibilidade e amizade.

Ao Professor Doutor Rui Oliveira pela disponibilidade e atenção sempre prestadas e pela supervisão deste trabalho.

Ao Professor Doutor Pereira Miguel, presidente do Instituto Nacional de Saúde Doutor Ricardo Jorge e ao Doutor João Lavinha, director do departamento de Genética por me terem dado a possibilidade de efectuar este trabalho neste instituto.

À Doutora Lúcia Lacerda, directora da unidade de Bioquímica-Genética do Centro de Genética Médica Doutor Jacinto de Magalhães – INSA, por me ter recebido e dado a oportunidade de efectuar este trabalho no laboratório que dirige.

À Doutora Maria Gil Ribeiro, directora da unidade de investigação e desenvolvimento do CGM – INSA pelos conselhos que me deu e pela simpatia e disponibilidade sempre demonstradas.

Ao Doutor Peter Jordan por me ter recebido no seu laboratório com toda a disponibilidade e simpatia e por todo o interesse que demonstrou pelo meu trabalho. Agradeço ainda aos elementos do seu grupo, Ana, Patrícia, Sónia, Vânia, Paulo e Bruno pela simpatia e amizade com que me receberam e por todos os

conhecimentos transmitidos. À Vânia uma palavra muito especial pois nada seria suficiente para agradecer toda a ajuda prestada durante a realização deste trabalho. Parte dos resultados aqui apresentados não teriam sido possíveis sem ti. Agradeço a tua paciência para me ensinar, a disponibilidade constante, os conselhos, o apoio e incentivo sempre que precisei. Por tudo isto e pela tua amizade o meu muito obrigada.

Ao Francisco, à Eugénia e à Isaura, técnicos da unidade de Bioquímica-Genética, pela forma simpática como sempre se disponibilizaram para me ajudar durante a execução deste trabalho.

Aos meus colegas do gabinete e grupo de investigação, Mia, Liliana, Carlos Bessa, Diogo, Joana e Carla Andreia (companheira na fase inicial), agradeço as palavras de apoio e todos os momentos de boa disposição e descontração proporcionados. À Carla Andreia o meu muito obrigada pela revisão desta tese.

A todos os amigos/as que acompanharam esta etapa de elaboração do trabalho prático, bem como a primeira fase do mestrado, em especial ao Carlos Melo e à Helena, o meu obrigada por todos os bons momentos que passamos juntos, pelo companheirismo, apoio, incentivo, dedicação e amizade.

À Sofia, Francisca, Natália, Nádia, Nanda, Carlos e Rui, cada um com uma amizade particular e especial, por me acompanharem e apoiarem com dedicação, atenção e interesse, por estarem sempre presentes e me fazerem sorrir nos bons e maus momentos, o meu obrigada especial...

A toda a minha família, em especial às minhas primas Romy e Isabel, obrigada pelo carinho, pelo apoio e por me fazerem sorrir, sempre presentes.

Aos meus pais, a quem devo tudo o que sou... pelo apoio e carinho incondicionais, por estarem sempre comigo, pela confiança que sempre depositaram em mim e no meu trabalho e empenho e por me ajudarem a realizar mais esta etapa do meu percurso académico e profissional, o meu mais sincero obrigada...

Resumo

O *splicing* do pré-mRNA é uma etapa essencial na expressão dos genes eucarióticos. Sabe-se actualmente que entre 15% a 60% das mutações exónicas ou intrónicas envolvidas em doença provocam alterações no processo de *splicing*.

Neste trabalho, para efectuar a análise funcional da mutação c.257C>T no processo de *splicing* do exão 3 do gene *IDS*, foi utilizada uma amostra de DNA genómico de um doente diagnosticado com Mucopolissacaridose tipo II.

Uma análise prévia do cDNA deste doente permitiu observar a ocorrência de dois transcritos na presença desta mutação. Um deles apresentou-se como um produto normal, apenas com a substituição C>T e origina uma proteína com a substituição aminoacídica P86L. No segundo transcrito observou-se a deleção dos primeiros 44 nucleótidos do exão 3, dado que, em vez do 3'ss normal foi utilizado um local críptico de *splicing* no interior do exão 3. Este local possui um valor de *score* mais elevado do que o local de *splicing* normal, sugerindo que a utilização deste estará possivelmente dependente de elementos *cis-acting* e/ou factores *trans-acting*. De facto, as análises bioinformáticas efectuadas previram alterações em motivos de ligação para as proteínas ASF/SF2, SC35 e hnRNP E1 e E2. Deste modo, no sentido de identificar quais os elementos/factores envolvidos na regulação do *splicing* desta região do exão 3 do gene *IDS* foi efectuada a construção de um minigene através da clonagem de um fragmento de gDNA do doente contendo parte do intrão 1, o exão 2, o intrão 2, o exão 3 e parte do intrão 3. O estudo *in vivo* deste minigene em células DT-1 permitiu verificar que o *splicing* ocorreu da mesma forma que no doente, sugerindo que este constructo é um sistema válido para futuramente se efectuar o estudo dos elementos *cis-acting* e factores *trans-acting* envolvidos na regulação do *splicing* desta região do gene *IDS*.

Assim, a validação deste minigene permite concluir que a técnica de clonagem e construção de minigenes foi implementada com sucesso no grupo de investigação em doenças lisossomais de sobrecarga do CGM-INSA, abrindo ainda a possibilidade de continuar a efectuar estudos funcionais que contribuam para o conhecimento dos mecanismos de *splicing* e de delinear novas abordagens terapêuticas nesta área.

Abstract

Pre-mRNA splicing is a critical step for eukaryotic gene expression. Actually it is known that 15% to 60% of disease-causing exonic or intronic mutations affect splicing.

In this study, to perform splicing functional analysis of c.257C>T mutation at exon 3 of *IDS* gene, a genomic DNA sample from a Mucopolysaccharidose II patient was used.

A preliminary analysis from patient cDNA allowed us to observe two transcripts. The first one showed only the C>T change which gives rise to a protein with the aminoacidic substitution P86L. The second one presented the deletion of the first 44 nucleotides of exon 3. This deletion was caused by the mentioned mutation which activates the use of an alternative downstream cryptic splice-site. This splice-site showed a higher score value than the constitutive splice-site, suggesting that this region is possibly under regulation of splicing *cis-acting* elements and/or *trans-acting* factors. In fact, bioinformatic analysis predicted the mutation to change ASF/SF2, SC35 and hnRNP E1 and E2 splicing factors binding sites. So, in order to identify the elements/factors involved in *IDS* gene exon 3 splicing regulation, a minigene was constructed with the interest gDNA fragment enclosing part of intron 1, exon 2, intron 2, exon 3 and part of intron 3. The minigene *in vivo* splicing analysis performed in a DT-1 cell line has shown the same splicing pattern observed in the patient, suggesting that this construct is a valid system for future studies of *cis-acting* elements and *trans-acting* factors implicated in *IDS* gene exon 3 splicing regulation.

As a result, minigene validation allow to conclude that cloning and minigene construction technique was successfully implemented in CGM-INSA lysosomal storage disorders research group, opening therefore the potential for further functional studies which can contribute to improve knowledge of general splicing mechanisms and to design new therapeutic strategies in this area.

Índice

AGRADECIMENTOS	vii
RESUMO	ix
ABSTRACT	x
LISTA DE ABREVIATURAS	xvi
CAPÍTULO I	1
INTRODUÇÃO	3
1.1 PRINCÍPIOS BÁSICOS DO PROCESSO DE <i>SPLICING</i>	3
1.1.1 O Reconhecimento do exão – um processo regulado por múltiplos factores.....	5
1.1.1.1 Estrutura exão-intrão - importância na selecção dos <i>splice-sites</i>	6
1.1.1.2 Elementos <i>cis-acting</i> e <i>trans-acting</i> envolvidos na regulação do <i>splicing</i>	8
1.1.1.3 O envolvimento da estrutura secundária do RNA e da síntese do pré-mRNA no processo de <i>splicing</i>	10
1.1.2 <i>Splicing</i> alternativo – um gene, vários transcritos, diferentes proteínas	12
1.1.3 Desregulação do processo de <i>splicing</i> e doença	14
1.1.3.1 Mutações de <i>Splicing</i> e Doenças Lisossomais de Sobrecarga	17
1.1.4 O uso de minigenes como método de estudo do processo de <i>splicing</i>	19
1.1.4.1 Construção de minigenes	20
1.1.4.2 Transfecção e obtenção de RNA	22
1.1.5 Utilização de minigenes para o estudo de mutações de <i>splicing</i> – alguns exemplos	23
CAPÍTULO II	29
2. OBJECTIVOS.....	31
CAPÍTULO III	33
3. MATERIAL E MÉTODOS	35
3.1 MATERIAL BIOLÓGICO.....	35
3.2 METODOLOGIA EXPERIMENTAL.....	35
3.2.1 Cultura celular de fibroblastos e obtenção de <i>pellets</i> celulares	35
3.2.2 Extracção de DNA genómico	36
3.2.3 Extracção de RNA total.....	36
3.2.4 Análise mutacional do cDNA do gene <i>IDS</i>	37
3.2.4.1 Síntese de cDNA	37
3.2.4.2 Amplificação por PCR dos fragmentos de cDNA <i>IDS</i>	38
3.2.4.3 Análise da reacção de PCR e purificação dos produtos amplificados	39
3.2.4.4 Sequenciação do cDNA <i>IDS</i>	39
3.2.5 Análise Bioinformática	41
3.2.6 Análise funcional do impacto da mutação c.257C>T no processo de <i>splicing</i> do gene <i>IDS</i> – Clonagem e construção de minigenes repórter	41

3.2.6.1 Amplificação por PCR dos fragmentos de DNA genómico de interesse	42
3.2.6.2 Análise da reacção de PCR e Sequenciação	43
3.2.6.3 Construção de minigenes repórter intermédios	44
3.2.6.3.1 Sub-clonagem dos fragmentos de interesse obtidos por PCR no vector pCR2.1®-TOPO®	44
3.2.6.3.2 Transformação de células quimicamente competentes.....	45
3.2.6.3.3 Obtenção de minigenes intermédios construídos no vector pCR2.1®- TOPO®	46
3.2.6.4 Construção de minigenes repórter finais.....	48
3.2.6.5 Cultura de células para transfecção	52
3.2.6.5.1 Transfecção	53
3.2.6.5.2 Extracção de RNA total e síntese de cDNA após transfecção	54
3.2.6.5.3 Amplificação e sequenciação dos fragmentos de cDNA do minigene <i>IDS</i> obtidos após transfecção	54
CAPÍTULO IV.....	57
4. RESULTADOS E DISCUSSÃO	59
4.1 IDENTIFICAÇÃO E CARACTERIZAÇÃO DA MUTAÇÃO C.257C>T NO GENE <i>IDS</i> DE UM DOENTE COM MUCOPOLISSACARIDOSE II.....	59
4.2 ANÁLISE BIOINFORMÁTICA	60
4.3 ANÁLISE FUNCIONAL DO IMPACTO DA MUTAÇÃO C.257C>T	67
4.3.1 Obtenção de minigenes repórter intermédios	67
4.3.2 Obtenção do minigene repórter final construído com a sequência mutada (c.257C>T) do gene <i>IDS</i>	70
4.3.3 Transfecção transiente do minigene repórter final (pcDNA3-Myc1® + fragmento gDNA <i>IDS</i> doente).....	73
CAPÍTULO V.....	75
5. CONCLUSÃO E PERSPECTIVAS FUTURAS	77
CAPÍTULO VI.....	79
6. BIBLIOGRAFIA	81

Índice de Figuras

Figura 1: A) Representação esquemática dos locais de <i>splicing</i> clássicos situados próximo das extremidades do intrão (5' <i>splice-site</i> ; <i>Branch site</i> ; <i>Polypyrimidine tract</i> (Y) _n ; 3' <i>splice-site</i>). Os motivos que flanqueiam os nucleótidos conservados destes locais clássicos são mostrados em baixo. Os nucleótidos que fazem parte das sequências consenso clássicas encontram-se a azul, excepto para o <i>branch point</i> A que é representado pela cor laranja. Adaptado de Cartegni <i>et al.</i> , (2002). B) Esquema representativo do processo de montagem do spliceossoma (este processo encontra-se descrito no texto). Adaptado de Query (2009).	4
Figura 2: Representação dos vários elementos que regulam o processo de <i>splicing</i> . Para além dos locais de <i>splicing</i> clássicos e de alguns elementos do spliceossoma, encontram-se representados os elementos <i>cis-acting</i> auxiliares (<i>Exonic splicing enhancers</i> e <i>silencers</i> – ESEs e ESSs; <i>Intronic splicing enhancers</i> e <i>silencers</i> – ISEs e ISSs), assim como os factores <i>trans-acting</i> - proteínas SR e hnRNP que interagem respectivamente com os elementos <i>cis-acting enhancers</i> e <i>silencers</i> . Adaptado de Pagani & Baralle (2004).	9
Figura 3: Diferentes modelos propostos para o processo de <i>splicing</i> alternativo (cada um destes modelos encontra-se explicado no texto). Adaptado de Matlin <i>et al.</i> (2005).	14
Figura 4: Esquema representativo da construção de um minigene num vector plasmídico. CMV – Citomegalovirus <i>transcriptional enhancer/promotor</i> ; RS#1 e RS#2 – locais de restrição situados no plasmídeo. Adaptado de Cooper (2005).	21
Figura 5: Esquema representativo dos locais de emparelhamento dos <i>primers</i> desenhados para amplificação dos fragmentos mutante e <i>wildtype</i> do gene <i>IDS</i> , assim como os locais de inserção nos vectores pCR2.1®-TOPO® e pcDNA3-Myc1®.	43
Figura 6: Vector plasmídico pCR2.1®-TOPO®	44
Figura 7: Esquema ilustrativo do processo de transformação do plasmídeo pCR2.1®-TOPO® em bactérias quimicamente competentes de <i>Escherichia coli</i>	46
Figura 8: Vector plasmídico pcDNA3-Myc1®	49
Figura 9: Esquema representativo da reacção de hidrólise do fragmento do gene <i>IDS</i> (doente) inserido no vector pCR2.1®-TOPO® e do vector pcDNA3-Myc1®.	50
Figura 10: Esquema ilustrativo do processo de transformação do plasmídeo pcDNA3-Myc1® em bactérias quimicamente competentes de <i>Escherichia coli</i>	51
Figura 11: Análise do cDNA do doente com a mutação c.257C>T no gene <i>IDS</i> . C – cDNA do indivíduo controlo; D – cDNA do doente; M – marcador de pesos moleculares.	59
Figura 12: Representação esquemática dos transcritos obtidos na presença da mutação c.257C>T. Encontram-se também representados os diferentes valores de <i>score</i> para o 5' <i>splice-site</i> (ss), 3' <i>splice-site</i> (ss) e 3' <i>splice-site</i> (ss) alternativo.	60
Figura 13: Representação das previsões efectuadas no <i>software Splicing Rainbow</i> para a sequência <i>wild-type</i> (A) e com a substituição nucleotídica C>T no exão 3 do gene <i>IDS</i> (B).	62

Figura 14: Representação das previsões efectuadas no <i>software ESE finder 3.0</i> . Previsão efectuada para a proteína SR – ASF/SF2; posição exónica, motivo e <i>score</i> para a sequência <i>wild-type</i> e com a mutação c.257C>T.	63
Figura 15: Representação das previsões efectuadas no <i>software ESE finder 3.0</i> . Previsão efectuada para a proteína SR – SC35; posição exónica, motivo e <i>score</i> para a sequência <i>wild-type</i> e com a mutação c.257C>T.	64
Figura 16: Modelo proposto para a regulação do <i>splicing</i> do exão 3 do gene <i>IDS</i> com base nas previsões efectuadas pelo <i>software ESE finder</i> (A) e <i>Splicing Rainbow</i> (B).	65
Figura 17: Amplificação em gradiente dos fragmentos de interesse do gene <i>IDS</i> para o controlo e doente. gDNA (DNA genómico) Controlo; gDNA (DNA genómico) Doente; $\Delta 57^{\circ}$ - 63° (Intervalo de temperaturas de amplificação - 57°C , 59°C , 61°C e 63°C); B – branco; M – marcador de pesos moleculares.	68
Figura 18: Selecção por PCR de colónias positivas (plasmídeos pCR2.1®-TOPO® [Invitrogen™] com fragmentos de interesse inseridos) para o controlo (A) e doente (B). Cl – Colónia; B – branco; M – marcador de pesos moleculares.	68
Figura 19: Identificação das substituições nucleotídicas ocorridas na sequência de gDNA do fragmento controlo por sequenciação automática. (A) Sequência parcial do gDNA clonado que permite identificar a substituição A>G no exão 2 do gene <i>IDS</i> . (B) Sequência parcial do gDNA clonado onde se pode observar a substituição T>C no intrão 2 do gene <i>IDS</i>	69
Figura 20: Digestão enzimática do vector pcDNA3-Myc1® e do minigene repórter intermédio contendo clonado o fragmento de gDNA <i>IDS</i> do doente com as enzimas de restrição <i>Bam</i> HI® e <i>Xho</i> I®. (A) Linha 1 - vector pcDNA3-Myc1® digerido; Linha 2 - vector pcDNA3-Myc1® não digerido. (B) Linha 1 – digestão do minigene intermédio; Seta vermelha – fragmento de interesse digerido; Seta verde – vector pcDNA3-Myc1® digerido. M – marcador de pesos moleculares.	71
Figura 21: Selecção por PCR de colónias positivas (pcDNA3-Myc1® [Invitrogen™] com o fragmento de interesse do doente inserido). Colónias positivas – número 11 e 12; B – branco; M – marcador de pesos moleculares.	71
Figura 22: Representação das sequências do gDNA <i>wild-type</i> e gDNA do doente em estudo com a mutação c.257C>T no gene <i>IDS</i> . (A) Sequência normal do fragmento de interesse do gene <i>IDS</i> obtida de um controlo. (B) Sequência do fragmento de interesse mutado pertencente ao doente, obtida após inserção no vector pcDNA3-Myc1®.	72
Figura 23: Representação esquemática do minigene repórter final construído com o fragmento de interesse do doente no vector plasmídico pcDNA3®-Myc. Encontram-se representados os <i>primers</i> (K-myc-Bam <i>forward</i> e “IDS3” <i>reverse</i>) utilizados para amplificação do fragmento após ocorrência do <i>splicing in vivo</i>	73
Figura 24: Amplificação do cDNA do fragmento de interesse do gene <i>IDS</i> do doente, obtido após ocorrência do processo de <i>splicing</i> no minigene repórter final transfectado em células DT-1. As setas cor de laranja indicam a presença dos dois transcritos. M – marcador de pesos moleculares; B – branco; A – amostra.	74

Índice de Tabelas

Tabela I. Sequência dos <i>primers</i> utilizados na amplificação dos fragmentos cDNA <i>IDS</i>	39
Tabela II. Mistura de reacção utilizada no PCR de sequenciação.	40
Tabela III. Condições do PCR de sequenciação.	40
Tabela IV. Sequência dos <i>primers</i> utilizados na amplificação dos fragmentos de gDNA <i>IDS</i> (normal e com a mutação c. 257C>T) e respectivas condições de amplificação por PCR.	43
Tabela V. Condições utilizadas na clonagem dos fragmentos de interesse no vector pCR2.1®-TOPO® [Invitrogen™].....	45
Tabela VI. Etapas seguidas no processo de clonagem nos plasmídeos pCR2.1®-TOPO® e pcDNA3-Myc1®.....	52
Tabela VII. Condições e quantidades utilizadas na transfecção em células DT-1 do plasmídeo pcDNA3-Myc1® construído com a sequência mutada (c.257C>T) do gene <i>IDS</i>	54
Tabela VIII. Sequência dos <i>primers</i> utilizados na amplificação dos fragmentos de cDNA <i>IDS</i> obtidos após transfecção e respectivas condições de amplificação por PCR.	55

Lista de abreviaturas

Abs – Absorvância

AMOs – *Antisense morpholino oligonucleotides*

Amp – Ampicilina

APC – *Adenomatous polyposis coli*

ATP – *Adenosine triphosphate*

BMD – Distrofia muscular de Becker

BPS – *Branch-point sequence*

BSA – *Bovine seric albumin*

cDNA – cadeia de DNA complementar ao mRNA

cSNPs – *Coding single-nucleotide polymorphisms*

CFTR – *Cystic fibrosis transmembrane conductance regulator*

CLN – Ceróide lipofuscinose neuronal

CTD – Domínio C-terminal

ddNTPs – Di-Desoxirribonucleótidos trifosfato

dNTPs – Desoxirribonucleótido trifosfato

DMD – Distrofia muscular de Duchenne

DMEM – *Dulbecco's modified eagle medium*

DNA – Ácido Desoxirribonucleico

Dscam – *Down syndrome cell-adhesion molecule*

DLS – Doença lisossomal de sobrecarga

DTT – 1,4-Dithiothreitol

EDTA – *Ethylenediamine tetra-acetic acid*

ESE – *Exonic splicing enhancer*

ESS – *Exonic splicing silencer*

FAP – *Familial adenomatous polyposis*

FBS – *Fetal bovine serum*

gDNA – DNA genómico

hnRNP – *heteronuclear ribonuclear protein*

IDS – *Iduronato-2-sulfatase*

ISE – *Intronic splicing enhancer*

ISS – *Intronic splicing silencer*

LB – *Luria Bertani*

LF2K – *Lipofectamina™ 2000*

mRNA – RNA mensageiro

NMD – *Nonsense-mediated mRNA-decay*

pb – Pares de bases

pol II – Polimerase II

PBS – *Phosphate buffer saline*

PCR – *Polymerase chain reaction*

PMM2 – *Phosphomannomutase 2*

PPT – *Polypyrimidine tract*

Pré-mRNA – Pré RNA mensageiro

rpm – Rotações por minuto

RNA – Ácido Ribonucleico

RNAi – RNA de interferência

RT-PCR – *Reverse transcriptase – Polymerase chain reaction*

snRNP – *small nuclear Ribonucleoprotein*

ss – *Splice-site*

SA – *Splicing alternativo*

SDS – *Sodium dodecil sulfate*

SF1 – *Splicing factor 1*

SMA – *Spinal muscular atrophy*

SMN1 – *Survival of motor neuron 1*

SMN2 – *Survival of motor neuron 2*

SOC – *Super optimal broth with catabolite repression*

SR – *Serine-Arginine rich*

TAE – Tampão *Tris*-acetato

Tris-HCl – *Tris*-cloreto de hidrogénio

TE – *Tris-EDTA*

UV-Vis – Ultravioleta-Visível

X-gal – 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside

Capítulo I

Introdução

Introdução

1.1 Princípios básicos do processo de *splicing*

Como material genético, a molécula de DNA oferece um modelo que orienta todas as actividades celulares e especifica o plano de desenvolvimento de todos os organismos multicelulares. Deste modo, o estudo da estrutura e função dos genes é fundamental para a compreensão da biologia molecular das células (Cooper & Hausman, 2009).

A expressão dos genes eucarióticos é um processo com várias etapas que inclui a transcrição do gene, o *splicing* e a poliadenilação do transcrito primário, assim como o transporte do RNA mensageiro (mRNA) para o citoplasma (Singer & Green, 1997), onde é traduzido em proteína. O *splicing* do pré-RNA mensageiro (pré-mRNA) foi descoberto no final dos anos 70 quando foi demonstrado que os genes eucarióticos continham determinadas sequências, os intrões, que não se encontravam presentes no mRNA maduro (Berget *et al.*, 1977; Chow *et al.*, 1977). Estudos subsequentes demonstraram que os intrões são removidos e os exões ligados numa ordem correcta por um complexo macromolecular denominado spliceossoma, constituído por cinco pequenos RNAs - U1, U2, U4, U5 e U6 em que cada um funciona como um complexo RNA-proteína conhecido como *small nuclear ribonucleoprotein* (snRNP) e por um elevado número de outras proteínas *non-snRNP*. Para que o processo de *splicing* possa ocorrer eficientemente são necessários quatro locais presentes em todos os intrões, entre os quais o 5'*splice-site* (5'ss), a *branch point sequence* (BPS), o *polypyrimidine tract* (PPT) e o 3'*splice-site* (3'ss) (Figura 1A). A montagem do spliceossoma (Figura 1B) é iniciada pela formação do complexo E que envolve a associação do U1 snRNP ao 5'ss e o reconhecimento do 3'ss por parte do factor auxiliar heterodimérico U2AF65/35 que se liga respectivamente ao PPT e à sequência terminal do intrão - 3'ss - AG, assim como a ligação da proteína *splicing factor 1* (SF1) à região BPS. De seguida, o U2 snRNP é associado à BPS, numa ligação dependente de ATP, permitindo a formação do complexo A. O subsequente recrutamento do tri-snRNP U4/U6•U5 resulta na formação do complexo B, que é seguida por uma série de rearranjos estruturais,

incluindo a perda dos U4 e U1 snRNPs, levando à formação do complexo catalítico C. Este complexo uma vez formado, permite as duas reacções de transesterificação sequenciais necessárias à ligação dos exões e libertação do *lariat* intrónico (Hastings & Krainer, 2001; Nilsen 2003; Jurica & Moore, 2003; Schellenberg *et al.*, 2008).

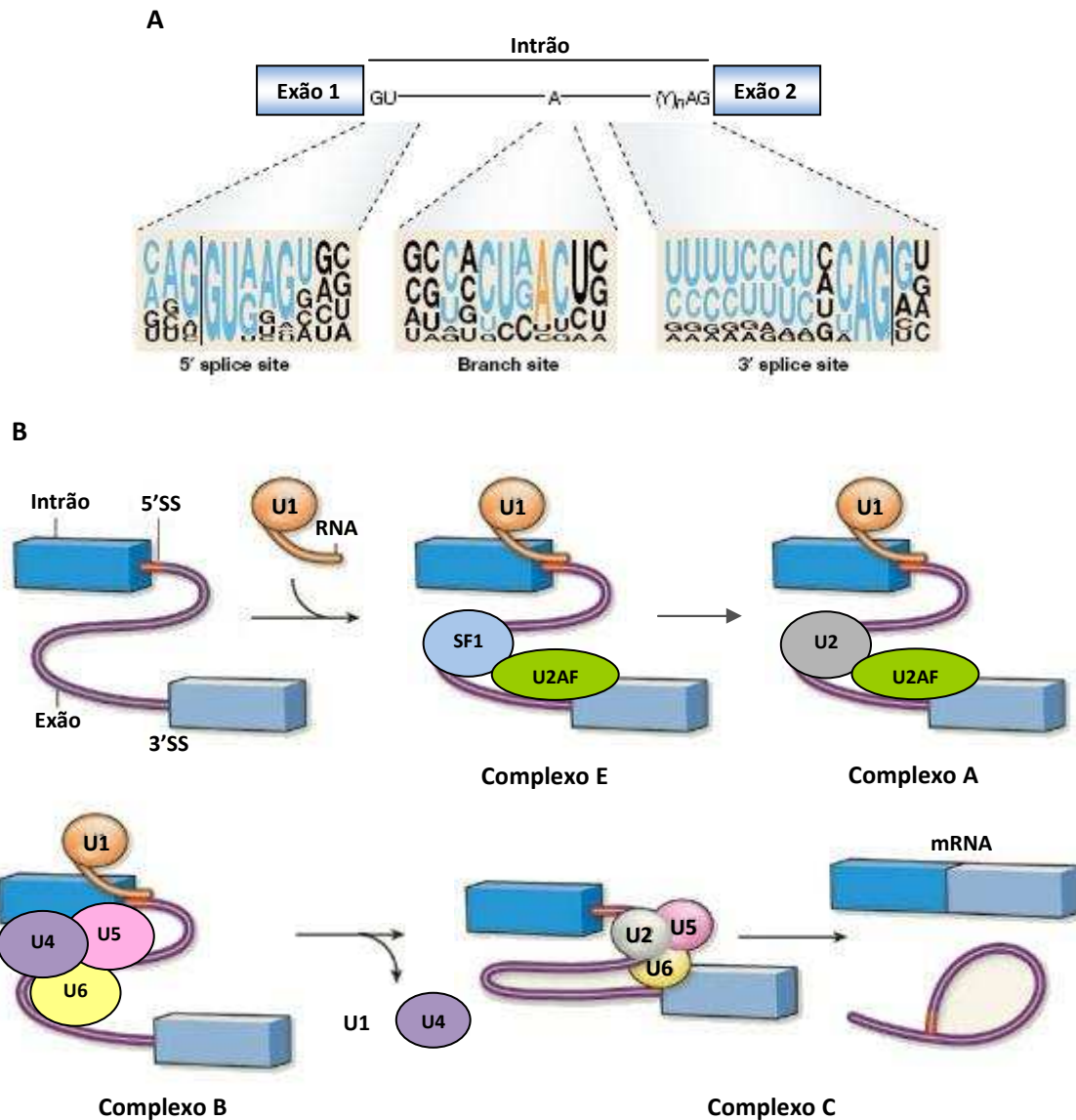


Figura 1: A) Representação esquemática dos locais de *splicing* clássicos situados próximo das extremidades do intrão (5'*splice-site*; *Branch site*; Polypyrimidine tract (Y)_n; 3'*splice-site*). Os motivos que flanqueiam os nucleótidos conservados destes locais clássicos são mostrados em baixo. Os nucleótidos que fazem parte das sequências consenso clássicas encontram-se a azul, excepto para o *branch point* A que é representado pela cor laranja. Adaptado de Cartegni *et al.*, (2002). B) Esquema representativo do processo de montagem do spliceossoma (este processo encontra-se descrito no texto). Adaptado de Query (2009).

Embora o spliceossoma catalise a quebra e ligação do RNA com grande exactidão, a flexibilidade inerente a este complexo enzimático faz com que este seja altamente sensível ao seu próprio mecanismo de regulação. Uma consequência frequente desta sensibilidade é a inclusão ou exclusão diferencial de exões no mRNA final num processo denominado *splicing* alternativo (SA) (Black 2003; House & Lynch, 2008). Estima-se que este processo ocorra em cerca de 60% dos genes humanos sendo um mecanismo pelo qual os organismos mais complexos regulam a expressão proteica, originando um proteoma diverso a partir de um genoma relativamente limitado (Black 2003; Stamm *et al.*, 2005; Ben-Dov *et al.*, 2008).

1.1.1 O Reconhecimento do exão – um processo regulado por múltiplos factores

O mecanismo de *splicing* tem vindo a ser estudado com grande detalhe. Contudo, ainda não se encontra totalmente esclarecido como é efectuada a selecção dos *splice-sites*. As sequências reguladoras deste processo (5' e 3'ss, BPS e PPT) apesar de conservadas são muito degeneradas, o que constitui o maior problema para a compreensão desta escolha, não sendo portanto ainda possível prever de um modo preciso quais os padrões de *splicing* de uma determinada sequência genómica. Deste modo, actualmente pensa-se que o reconhecimento e selecção eficientes dos *splice-sites* de um determinado pré-mRNA, pelo spliceossoma, com o consequente processamento dos exões sejam mediados por diversos factores (Stamm *et al.*, 2005; Hertel 2008). Entre estes encontram-se a estrutura exão-intrão, a “força” dos *splice-sites*, a presença ou ausência de sequências reguladoras do *splicing* (*enhancers* – activadores e *silencers* – repressores), a presença ou ausência de estruturas secundárias no RNA, assim como o processo de síntese do pré-mRNA pela RNA polimerase II (pol II). Cada um destes componentes contribui para a afinidade global do spliceossoma a cada exão, determinando assim o seu nível de inclusão no RNA maduro (Matlin *et al.*, 2005; House & Lynch 2008; Hertel 2008).

1.1.1.1 Estrutura exão-intrão - importância na selecção dos *splice-sites*

Tipicamente um gene humano contém exões relativamente pequenos (50-250 pares de bases de comprimento) separados por intrões maiores (centenas a milhares de pares de bases) que em média são responsáveis por mais de 90% da constituição do transcrito primário (Wang & Burge, 2008). O reconhecimento e emparelhamento preciso do spliceossoma nos *splice-sites* 5' e 3' é necessário para que os intrões possam ser removidos e os exões ligados correctamente. O local de *splicing* a 5' nos eucariotas superiores compreende 9 nucleótidos, ao passo que o 3'ss é definido pelos três elementos - BPS, PPT e a própria sequência da junção 3'exão-intrão, que se localizam até quarenta nucleótidos a montante desta junção (Horowitz & Krainer, 1994; Chiara & Reed, 1995). A classificação destes *splice-sites* é efectuada de acordo com o grau de complementaridade entre o U1 snRNP e o 5'ss e pela extensão do PPT ao qual se liga o factor U2AF. Deste modo, quanto maior a complementaridade com o 5'ss e quanto mais longos os PPTs, maior será a afinidade entre os *splice-sites* e os componentes do spliceossoma traduzindo-se num reconhecimento mais eficiente dos exões (Hertel 2008).

Por outro lado o tamanho quer dos exões quer dos intrões parece estar também relacionado com o reconhecimento e selecção dos *splice-sites* (Berget 1995). Por exemplo, o aumento do tamanho dos exões, no caso de mamíferos, resultou na ocorrência do *skipping* desses mesmos exões. Contudo, os mesmos exões foram incluídos no transcrito final quando os intrões flanqueantes eram pequenos. Desta forma, quando os exões ou intrões são de pequeno tamanho, o reconhecimento dos *splice-sites* é mais eficiente, sugerindo que estes são também identificados através de um mecanismo que reconhece e avalia o comprimento nucleotídico óptimo e que este influencia significativamente, tanto a eficiência do *splicing* do pré-mRNA como a escolha de *splice-sites* alternativos (Hertel 2008). Esta é uma hipótese importante, dada a distribuição divergente do comprimento dos intrões no genoma humano e a hipótese de que o spliceossoma utiliza dois modelos de reconhecimento para definir os *splice-sites*. Um destes modelos designado "intron definition" estabelece que os 5' e 3' *splice-sites* dos intrões são

directamente identificados pela maquinaria de *splicing* como uma unidade. Por outro lado, um modelo denominado “exon definition” define que os *splice-sites* que flanqueiam o exão são inicialmente reconhecidos e subsequentemente emparelhados pelos elementos do spliceossoma (Berget 1995). Porém, apesar de estes dois modelos ainda não se encontrarem provados, algumas análises recentes da co-evolução dos locais de *splicing* a 5’ e a 3’ têm detectado interacções predominantes “across-exon” em humanos e em ratos, e interacções do tipo “across-intron” em invertebrados, plantas e fungos, corroborando uma preferência do modelo “exon definition” em mamíferos e “intron definition” na maioria dos outros metazoários (Xiao *et al.*, 2007).

Num outro estudo recente (Fox-Walsh *et al.*, 2005), foi demonstrado que a preferência pelos *splice-sites* “across-introns” decresceu quando o tamanho do intrão se situava entre 200 e 250 nucleótidos. Por outro lado, quando o tamanho do intrão se encontrava acima deste valor, foram reconhecidos os *splice-sites* “across-exons”. Verificaram ainda que a selecção dos *splice-sites* “across introns” foi significativamente mais eficiente do que quando efectuada “across-exons”, na qual se verifica uma elevada inclusão de exões com *splice-sites* mais “fracos”, favorecendo a ocorrência de *splicing* alternativo. Este estudo demonstrou assim que o tamanho dos intrões influencia a probabilidade com que um exão sofre *splicing* alternativo. Estas observações permitem perceber que a arquitectura exão-intrão define, tanto os mecanismos de reconhecimento dos *splice-sites* como influencia a frequência com que ocorre o *splicing* alternativo do pré-mRNA.

Com o aumento do conhecimento nesta área tem sido obtida informação acerca da estrutura exão-intrão, permitindo o desenvolvimento de modelos estatísticos que auxiliem a previsão da localização dos motivos dos 5’ e 3’ *splice-sites* de uma forma mais precisa. Algumas ferramentas para o cálculo dos *scores* das sequências consenso dos *splice-sites* encontram-se disponíveis na internet, como por exemplo: *MaxEntScan* (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html) (Yeo & Burge, 2004), *Splice Site Prediction by Neural Network* (http://www.fruitfly.org/seq_tools/splice.html) (Reese *et al.*, 1997), *Human Splicing Finder* (<http://www.umd.be/HSF/>) (Hubbard *et al.*, 2007) e GENSCAN (<http://genes.mit.edu/GENSCAN.html>) (Burge & Karlin, 1997).

1.1.1.2 Elementos *cis-acting* e *trans-acting* envolvidos na regulação do *splicing*

Apesar do processo de *splicing* possuir um enorme potencial para ocorrer de forma errada, parece suceder com enorme exactidão, em parte devido ao envolvimento de sequências *cis-acting* auxiliares existentes no transcrito primário, que para além dos *splice-sites* influenciam a selecção destes, activando ou reprimindo a sua utilização (Ladd & Cooper, 2002; Wang & Burge, 2008; Matlin *et al.*, 2005). Estas sequências *cis-acting* são convencionalmente designadas como *Exonic Splicing Enhancers* (ESEs) ou *Exonic Splicing Silencers* (ESSs) se respectivamente, activam ou reprimem a inclusão do exão onde a sua sequência se encontra, ou como *Intronic Splicing Enhancers* (ISEs) ou *Intronic Splicing Silencers* (ISSs) se respectivamente, activam ou reprimem o uso de *splice-sites*/exões adjacentes à sua localização intrónica (Blencowe 2000; Cartegni *et al.*, 2002; Wang & Burge, 2008; Hertel 2008). De um modo geral, estes elementos reguladores têm como função recrutar factores de *splicing trans-acting* que por vários mecanismos, activam ou reprimem o reconhecimento de *splice-sites* ou a montagem do spliceossoma (Ladd & Cooper, 2002; Cartegni *et al.*, 2002; Matlin *et al.*, 2005; Wang & Burge, 2008).

Os ESEs, são normalmente reconhecidos por factores *trans-acting* pertencentes a uma família de proteínas conhecida como proteínas SR, que contêm na sua estrutura um domínio de ligação ao RNA, assim como um domínio carboxi-terminal rico em dipéptidos arginina-serina (Domínio RS) (Cartegni *et al.*, 2002; Black 2003; Matlin *et al.*, 2005; Long & Cáceres, 2009). Porém, a regulação do *splicing* do pré-mRNA depende também de outros elementos já mencionados, os ESSs, ISEs e ISSs (Figura 2). O mecanismo de acção dos elementos envolvidos no silenciamento não se encontra ainda totalmente esclarecido. Contudo, parecem interagir com repressores pertencentes à família de proteínas *heterogeneous nuclear ribonuclearprotein* (hnRNP). Estas proteínas, de um modo similar às proteínas SR, possuem na sua estrutura um domínio de ligação ao RNA, associado a

um domínio auxiliar normalmente envolvido em interações proteína-proteína (Cartegni *et al.*, 2002; Hertel 2008).

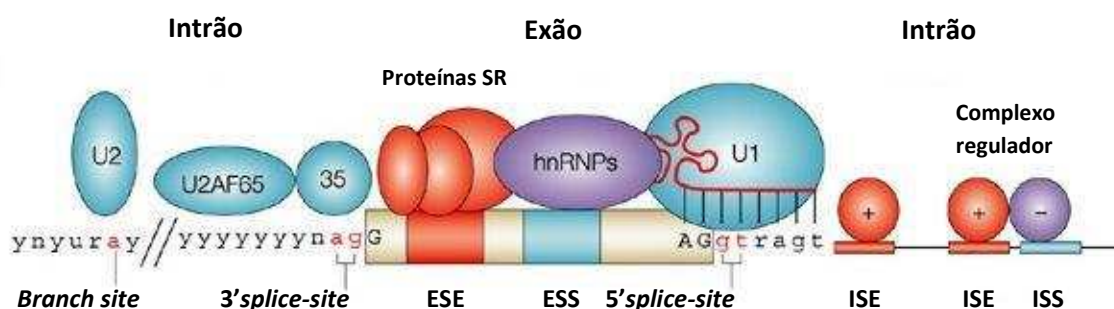


Figura 2: Representação dos vários elementos que regulam o processo de *splicing*. Para além dos locais de *splicing* clássicos e de alguns elementos do spliceossoma, encontram-se representados os elementos *cis-acting* auxiliares (*Exonic splicing enhancers* e *silencers* – ESEs e ESSs; *Intronic splicing enhancers* e *silencers* – ISEs e ISSs), assim como os factores *trans-acting* - proteínas SR e hnRNP que interagem respectivamente com os elementos *cis-acting enhancers* e *silencers*. Adaptado de Pagani & Baralle (2004).

Estes elementos *cis-acting* reguladores do *splicing* (activadores e repressores) encontram-se presentes nas proximidades das junções exão-intrão, sugerindo que a sua acção combinada, activando ou reprimindo determinados locais de *splicing*, influencia a probabilidade de inclusão de um exão (House & Lynch, 2008; Hertel 2008). Alguns estudos (Zhu *et al.*, 2001; Singh *et al.*, 2004) corroboram este modelo, sugerindo que o reconhecimento de cada exão é influenciado por múltiplos elementos *cis-acting* distintos e que o *ratio* de algumas formas de *splicing* alternativo para um dado pré-mRNA resulta de um balanço entre uma regulação positiva e negativa efectuada por estes elementos. Globalmente, estes apresentam sequência variável, são relativamente pequenos, individualmente fracos e presentes em mais do que uma cópia. Apresentam-se ainda conservados entre espécies e possivelmente entre genes que obedecem ao mesmo modo de regulação. Contudo, as suas sequências são normalmente degeneradas colocando dificuldades na sua identificação (Ladd & Cooper, 2002; Cartegni *et al.*, 2002).

Para além de análises experimentais no âmbito da genética molecular, estudos bioinformáticos permitiram desenvolver *software* para prever as

sequências destes elementos reguladores. Alguns programas encontram-se disponíveis na internet: *ESEfinder* (<http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home>) (Cartegni *et al.*, 2003), *Splicing Rainbow* (<http://www.ebi.ac.uk/asd-srv/wb.cgi?method=8>) (Stamm *et al.*, 2006) e *PESX: Putative Exonic Splicing Enhancers/Silencers* (<http://cubweb.biology.columbia.edu/pesx/>) (Zhang & Chasin, 2004; Zhang *et al.*, 2005).

O elevado envolvimento destes elementos *cis-acting* no processo de *splicing* faz com que facilmente se tornem um alvo de mutações patogénicas (Cartegni *et al.*, 2002). No entanto, estas ferramentas informáticas ainda não permitem identificar estes elementos com precisão, sendo sempre necessário validar as suas previsões através de análises experimentais.

1.1.1.3 O envolvimento da estrutura secundária do RNA e da síntese do pré-mRNA no processo de *splicing*

A estrutura secundária do RNA tem vindo a ser cada vez mais reconhecida como um modificador dos eventos de *splicing* (Buratti & Baralle, 2004). A maioria dos factores proteicos que se ligam ao RNA interagem com a sua estrutura de cadeia simples. Contudo, as sequências alvo do RNA onde se ligam estes factores encontram-se habitualmente presentes num “loop” formado pela própria estrutura do RNA e não na estrutura de cadeia simples (Pagani & Baralle, 2004). Ao nível local, as conformações do RNA regulam o processo de *splicing* afectando os elementos básicos (*splice-sites* e outros elementos *cis-acting* – activadores e repressores que permitem a identificação de um exão, inibindo ou activando assim a montagem do spliceossoma (Buratti & Baralle, 2004; Hertel 2008).

Alguns trabalhos têm demonstrado a importância da estrutura secundária do RNA na selecção dos *splice-sites*. Por exemplo, um estudo no gene *tau* demonstrou que a ocorrência de uma mutação ao nível do 5'ss provocou alterações na estrutura secundária do RNA (numa região *stem loop*) rompendo assim as interacções com elementos do spliceossoma e consequentemente alterando o processo de *splicing* (Donahue *et al.*, 2006). Adicionalmente, análises computacionais demonstraram

que aproximadamente 5% dos eventos de *splicing* alternativo se correlacionam com a presença de estruturas secundárias estáveis (Johnson *et al.*, 2003). Estes exemplos apoiam a ideia de que as estruturas secundárias de RNA terão um papel mais significativo na regulação do reconhecimento dos *splice-sites* do que é actualmente considerado.

Algumas observações recentes indicam que o processo de *splicing* será regulado não apenas por todos os factores já descritos anteriormente, mas também por um mecanismo mais complexo envolvendo a maquinaria de transcrição. De facto a transcrição e o processamento do pré-mRNA não são eventos independentes. Pelo contrário, existe um elevado grau de coordenação tanto temporal como espacial uma vez que as três reacções envolvidas no processamento do pré-mRNA (*5'-capping*, *splicing* e *3'-polyadenylation*) ocorrem em íntima associação com o processo de alongamento pela RNA polimerase II (pol II) (Cáceres & Kornblihtt, 2002). Desta forma, algumas análises têm apoiado esta hipótese, demonstrando que o processo de reconhecimento dos *splice-sites* pode ocorrer co-transcripcionalmente, isto é, os *splice-sites* de um exão podem ser identificados pelo spliceossoma enquanto os exões situados a jusante aguardam ainda a sua síntese pela pol II (Kornblihtt 2005; Tardiff *et al.*, 2006). Estes efeitos da transcrição no processo de *splicing* podem ser explicados segundo dois modelos. Um dos modelos sugere que os factores de *splicing* se associam ao domínio C-terminal (CTD) da pol II quando esta se encontra próxima do promotor (Cramer *et al.*, 1997). Durante a transcrição, elementos constituintes deste domínio são fosforilados tornando a pol II processiva e com capacidade de interagir com os vários factores de *splicing* existentes. Contudo, diferenças na estrutura do promotor podem levar a diferenças nos factores de *splicing* que se associam ao domínio CTD da pol II e consequentemente influenciar as decisões da maquinaria de *splicing* (de la Mata & Kornblihtt, 2006). Um segundo modelo complementar propõe que a cinética de transcrição da pol II influencia o *splicing* alternativo, uma vez que os *splice-sites* dos exões alternativos terão menor probabilidade de ligação por parte da maquinaria de *splicing* do que os exões constitutivos, podendo assim necessitar de um maior período de tempo para interagir com o spliceossoma, e o seu uso ser apenas

favorecido quando a taxa de alongamento da pol II é mais baixa (de la Mata *et al.*, 2003).

As relativas contribuições de cada um dos parâmetros mencionados, permitem controlar o reconhecimento inicial dos exões e conseqüentemente a eficácia com que os seus *splice-sites* são identificados e os respectivos intrões flanqueantes removidos (Hertel 2008). Deste modo, o processo de *splicing* pode ser visto como um processo onde uma série de “checkpoints” permitem ao spliceossoma experimentar diferentes opções e ajustá-las de acordo com o que é ditado localmente e requerido por cada célula em específico. Esta flexibilidade e controlo permitem de certo modo, prever a extensão do processo de *splicing* alternativo, actualmente reconhecido como ubíquo e essencial para a diversidade funcional requerida pelos organismos eucarióticos mais complexos (House & Lynch, 2008).

1.1.2 *Splicing* alternativo – um gene, vários transcritos, diferentes proteínas

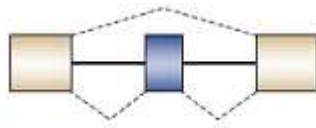
O *splicing* alternativo (SA) é um mecanismo fundamental que permite a síntese de diversas proteínas estrutural e funcionalmente diferentes a partir do mesmo gene. Estima-se que este processo ocorra em cerca de 60% dos genes humanos, tendo a capacidade de determinar a localização intracelular, actividade enzimática, estabilidade, propriedades de ligação e modificações pós-tradução de um vasto número de proteínas. Estas observações sugerem assim que o SA é um mecanismo responsável pela elevada complexidade do proteoma humano (Cáceres & Kornblihtt, 2002; Black 2003; Stamm *et al.*, 2005; Ben-Dov *et al.*, 2008).

Um gene humano tem a capacidade de gerar, em média, dois a três transcritos (Stamm *et al.*, 2005). Contudo, existem genes capazes de codificar dezenas a centenas de proteínas diferentes (Cartegni *et al.*, 2002). O gene *Dscam* (*Down syndrome cell-adhesion molecule*) de *Drosophila melanogaster* exemplifica a extrema diversidade estrutural que o processo de SA é capaz de gerar. Este gene codifica uma proteína de membrana envolvida na conectividade neuronal e o seu pré-mRNA pode ser processado de diversas formas, dando origem a 38 016

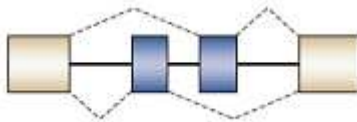
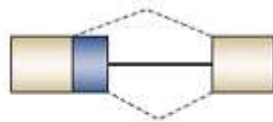
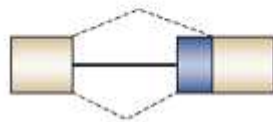
transcritos maduros diferentes, mais do que o dobro do número de genes existentes em todo o genoma de *Drosophila* (Schmucker *et al.*, 2000). As proporções das diferentes proteínas produzidas por SA podem variar de acordo com o tipo celular, o estado de desenvolvimento, processos fisiológicos, diferenciação sexual, apoptose e como resposta a factores de stress (Ladd & Cooper, 2002; Pajares *et al.*, 2007).

A regulação do *splicing* alternativo, permitindo a inclusão ou exclusão de determinados exões, pode ser conseguida através da modulação dos diversos componentes que interferem com o reconhecimento desses exões (elementos constituintes do spliceossoma; sequências *splice-site*; sequências dos elementos activadores e repressores; proteínas *trans-acting* reguladoras; estrutura exão-intrão; formação de estruturas secundárias de RNA e síntese do transcrito primário (pol II). As variações ocorridas nas actividades ou quantidades dos factores gerais de *splicing* e/ou reguladores de *splicing* específicos de um gene podem explicar os vários eventos de SA observados durante o desenvolvimento ou em diferentes tipos de células (Cáceres & Kornblihtt, 2002; Hertel 2008).

A produção de diversas isoformas a partir do mesmo transcrito por SA (Figura 3) é um acontecimento muito comum em células de mamíferos, que torna a já difícil tarefa de identificar correctamente os *splice-sites* ainda mais complexa. Um dos modelos mais comum de SA é o modelo *cassette exons*. Estes são exões que podem ser independentemente incluídos ou excluídos do mRNA final (Figura 3A). Outro tipo de SA é o modelo de exões mutuamente exclusivos que envolve a selecção de apenas um exão a partir de duas ou mais variantes exónicas existentes (Figura 3B). A selecção alternativa de um 5'ss ou de um 3'ss localizado no interior de um exão é também frequente, representando eventos de modificação exónica (Figura 3C) (Cartegni *et al.*, 2002; Matlin *et al.*, 2005; Corcos & Solier, 2005; Blencowe 2006). Os restantes modelos incluem a retenção de intrões (Figura 3D) e o SA combinado com o uso de promotores alternativos (Figura 3E) ou de locais de poliadenilação (poli A) alternativos (Figura 3F) (Matlin *et al.*, 2005; Corcos & Solier, 2005; Blencowe 2006).

A) *Cassette exons*

B) Exões mutuamente exclusivos

C) *Competing 5'splice-sites**Competing 3'splice-sites*

D) Retenção intrónica



E) Promotores múltiplos

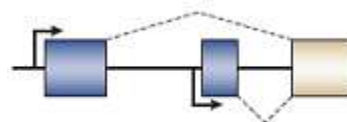
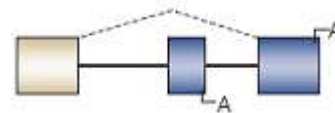
F) *Multiple poly(A) sites*

Figura 3: Diferentes modelos propostos para o processo de *splicing* alternativo (cada um destes modelos encontra-se explicado no texto). Adaptado de Matlin *et al.* (2005).

1.1.3 Desregulação do processo de *splicing* e doença

Nas últimas duas décadas, na área das doenças genéticas humanas, tem sido dada especial atenção à pesquisa de mutações que afectam o processo de *splicing* e para as quais têm vindo a ser formuladas e testadas novas terapias (Garcia-Blanco *et al.*, 2004; Cooper *et al.*, 2009).

Para além das doenças causadas por mutações exónicas, incluindo mutações *missense*, *nonsense* e *frame-shift* que originam proteínas aberrantes, e mutações em regiões não codificantes que afectam a eficiência da transcrição ou a estabilidade do mRNA, sabe-se actualmente que entre 15% a 60% das mutações exónicas ou intrónicas alteram o normal padrão de *splicing* (López-Bigas *et al.*, 2005; Wang & Cooper, 2007). Estas mutações, para além de poderem ser a causa

directa de doença, podem também modificar a sua gravidade fenotípica, assim como estarem ligadas a uma maior ou menor susceptibilidade de desenvolver doença (Wang & Cooper, 2007).

A maioria destas mutações pode afectar elementos *cis-acting* (junções exão-intrão, PPT, BPS e sequências exónicas/intrónicas activadoras ou repressoras) assim como factores *trans-acting* necessários à regulação do processo de *splicing*. No caso das mutações *cis-acting* tanto o uso de *splice-sites* constitutivos como alternativos pode ser afectado. A ruptura do *splicing* constitutivo resulta na maioria das vezes na perda da expressão génica, ao passo que uma mutação *cis-acting* que active ou inactive um de dois locais de *splicing* alternativos irá forçar a expressão de um desses padrões de SA. Deste modo, embora seja expresso um mRNA normal, a sua expressão inapropriada num tecido ou etapa do desenvolvimento poderá resultar em doença. Por outro lado, as mutações associadas a elementos *trans-acting* podem afectar o normal funcionamento da maquinaria basal de *splicing* assim como factores proteicos que intervêm no processo de SA. Alterações ocorridas nas estequiometrias relativas destes componentes na célula podem também causar anomalias no *splicing* e consequentemente doença. As mutações que afectam a maquinaria basal de *splicing* podem afectar o *splicing* de todos os pré-mRNAs, ao passo que as mutações que afectem um elemento regulador do SA irão apenas ter influência no conjunto de pré-mRNAs regulados por esse factor (Faustino & Cooper, 2003; Pagani & Baralle, 2004; Garcia-Blanco *et al.*, 2004; Wang & Cooper, 2007; Cooper *et al.*, 2009).

A maioria das mutações que alteram o processo normal de *splicing* são substituições nucleotídicas que ocorrem em elementos *cis-acting* contidos em segmentos exónicos ou intrónicos. Estas mutações podem causar o *skipping* de determinado exão, o uso de um local críptico de *splicing* a 5' ou 3', provocar a retenção de um intrão ou ainda criar um novo local de *splicing*. Na maior parte dos casos, o uso de *splice-sites* não naturais ou a retenção intrónica introduzem codões stop prematuros no mRNA activando um mecanismo de vigilância utilizado pela célula denominado *nonsense-mediated mRNA-decay* (NMD) que tipicamente degrada os transcritos mutados, fazendo assim com que o alelo mutado perca a sua

função (Faustino & Cooper, 2003; Pagani & Baralle, 2004; Garcia-Blanco *et al.*, 2004; Cooper *et al.*, 2009; Wang & Cooper, 2007).

A variação genética em polimorfismos nucleotídicos sinónimos localizados em regiões codificantes (*coding single-nucleotide polymorphisms* - cSNPs) é a principal causa das diferenças fenotípicas encontradas individualmente, incluindo aquelas que se tornam relevantes no aparecimento de uma doença (Cáceres & Kornblihtt, 2002; Garcia-Blanco *et al.*, 2004; Wang & Cooper, 2007; Cooper *et al.*, 2009). Por exemplo, para o gene *CFTR* (*Cystic fibrosis transmembrane conductance regulator*) um estudo envolvendo uma análise mutacional demonstrou que cerca de um quarto das substituições nucleotídicas encontradas nos exões 9 e 12 apesar de sinónimas (codificam para o mesmo aminoácido) alteravam a eficiência do processo de *splicing* (Pagani *et al.*, 2005). Deste modo, os cSNPs possuem um peso importante como modificadores da gravidade fenotípica e susceptibilidade de um indivíduo para uma determinada doença genética (Wang & Cooper, 2007; Cooper *et al.*, 2009).

Para uma correcta interpretação das bases moleculares das doenças hereditárias, e para que polimorfismos neutrais não sejam enganosamente considerados como mutações, será necessário um estudo dos padrões de *splicing* dos cSNPs por cDNA ou por análise funcional. Por exemplo, a detecção de uma mutação *missense* pode levar a supor que o aminoácido afectado seja crítico para a funcionalidade da proteína em causa, quando de facto após uma análise mais cuidada se percebe que a real causa da doença envolve a quebra de um ESE que leva ao *skipping* de um exão ou a outras alterações de *splicing*. Estas considerações impõem assim, a necessidade de uma análise do transcriptoma para além da análise genómica habitual, com o objectivo de avaliar o verdadeiro impacto do processo de *splicing* nas doenças humanas (Cáceres & Kornblihtt, 2002).

1.1.3.1 Mutações de *Splicing* e Doenças Lisossomais de Sobrecarga

As mutações de *splicing*, que em alguns casos originam transcritos não funcionais, têm sido identificadas em muitos doentes diagnosticados com uma doença lisossomal de sobrecarga (DLS). As DLS são um grupo de doenças hereditárias do metabolismo, em que mais de 50 são conhecidas, causadas pela actividade deficiente de uma proteína lisossomal, resultando na acumulação intra-lisossomal de metabolitos não degradados (Ballabio & Gieselmann, 2009).

As DLS podem ser agrupadas segundo várias classificações, mas como defendem Futerman & van Meer (2004), a mais útil será baseada na caracterização da enzima ou proteína deficiente, em vez de se basear na natureza do substrato acumulado. Esta última terá sido a causa para a caracterização errada de doenças para as quais estava identificado o substrato acumulado, anos antes de se descobrir o seu defeito enzimático. Segundo estes autores as DLS podem ser divididas em esfingolipidoses, mucopolissacaridoses, oligossacaridoses, glicoproteinoses, doenças causadas por defeitos em proteínas integrais de membrana e ainda uma subclassificação denominada “outras” onde se incluem as patologias que não se inserem em nenhum dos grupos anteriores, como por exemplo a galactosialidose, a deficiência de múltiplas sulfatases, as mucolipidoses II e III e as ceróide lipofuscinoses neuronais (CLNs). O modo de transmissão das DLS é autossómico recessivo, à excepção da doença de *Fabry* e da Mucopolissacaridose II (Síndrome de *Hunter*), cujo modo de transmissão é heterossómico recessivo (Meikle & Hopwood, 2003). Embora as DLS sejam individualmente raras, no seu conjunto apresentam uma incidência de aproximadamente 1:7700 nascimentos (Meikle *et al.*, 2006), representando, por isso, um reconhecido problema de saúde a nível mundial. Recentemente, para Portugal, a prevalência destas doenças foi estimada em 1:4000 nascimentos (Pinto *et al.*, 2004). A maioria das DLS conhecidas são consideradas doenças monogénicas associadas a uma considerável variabilidade clinicopatológica e heterogeneidade alélica resultante da presença de diferentes tipos de mutações (*missense*, *nonsense*, *splice-site*, deleções parciais e inserções) num único gene, embora em geral para cada uma estejam também descritas numerosas mutações no mesmo gene (Futerman & van Meer, 2004).

Para um grupo de DLS relativamente frequentes (Fabry; Gaucher; Niemann-Pick tipo A, B e C; Gangliosidoses 1 e 2; Mucopolissacaridoses II, IIIA e IIIB, VI e Ceróide lipofuscinose neuronal tipo 2 e 3) a percentagem de mutações de *splicing* é em média cerca de 10% (The Human Gene Mutation Database - <http://www.hgmd.cf.ac.uk/ac/index.php>). Contudo, este valor poderá encontrar-se subestimado dado que muitas vezes apenas são efectuadas análises ao DNA genómico (gDNA) que não permitem clarificar qual o verdadeiro impacto de muitas mutações silenciosas, *missense* e *nonsense*.

No caso da doença lisossomal de sobrecarga Mucopolissacaridose tipo II (Síndrome de *Hunter*) causada por alterações no gene *IDS*, que codifica para a enzima iduronato-2-sulfatase, alguns estudos de caracterização molecular (Karsten *et al.*, 1998; Lualdi *et al.*, 2006; Alves *et al.*, 2006) têm descrito diferentes mutações que alteram o processo normal de *splicing*, como sendo a causa frequente da doença. Especificamente no estudo molecular efectuado por Alves *et al.* (2006) foi identificada no gene *IDS* a mutação estudada no presente trabalho: c.257C>T.

O gene *IDS* apresenta *splicing* alternativo com pelo menos oito transcritos diferentes conhecidos (Ensembl - http://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000010404). Este número elevado de opções de transcrição aponta para uma regulação complexa do processo de *splicing* neste gene, que provavelmente requer a interacção de vários elementos *cis-acting* auxiliares. Deste modo, um mecanismo com esta complexidade poderá fazer aumentar a incidência de mutações que originam a ruptura do normal processo de *splicing*, constituindo uma importante fonte de patogenicidade no locus *IDS*.

Parte do conhecimento sobre os mecanismos moleculares básicos subjacentes ao processo de *splicing* do pré-mRNA, tem sido obtido por estudos funcionais de mutações específicas em genes particulares, com o objectivo de perceber de que modo a sua presença afecta o *splicing* desses genes. Assim, tanto a mutação referida para o gene *IDS* como outras envolvidas na origem, tanto de doenças de sobrecarga lisossomal como de outras, representam uma importante oportunidade para efectuar estudos que permitam clarificar o processo de *splicing* em geral.

1.1.4 O uso de minigenes como método de estudo do processo de *splicing*

Dada a importância biológica e biomédica do processo de *splicing* têm surgido nos últimos anos um número elevado de estudos que visam os mecanismos que controlam as decisões deste processo assim como as consequências biológicas dos diferentes transcritos alternativos. Estas análises têm permitido o desenvolvimento de diversas metodologias de entre as quais se destaca o uso de sistemas de minigenes construídos para a análise dos elementos que intervêm neste processo e que permitem recapitular *in vitro* a regulação observada *in vivo* (Garcia-Blanco 2005).

Os constructos de minigenes são uma importante ferramenta para a identificação e análise *in vivo* tanto dos elementos reguladores *cis-acting* como dos factores *trans-acting* que determinam a eficiência e regulam o processo de *splicing*. A expressão dos pré-mRNAs em minigenes através da transfecção transitória permite efectuar uma análise rápida do ganho ou perda de função para os elementos *cis-acting* e os factores *trans-acting* que afectam a regulação do *splicing*.

O primeiro passo de uma investigação que tenha como objectivo analisar as características que regulam o *splicing* é estabelecer qual ou quais os pontos específicos a estudar. Assim, os minigenes são utilizados habitualmente para: (i) determinar o papel desempenhado pelos *splice-sites* no reconhecimento exónico, (ii) identificar os elementos exónicos e intrónicos que activam ou reprimem o processo de *splicing*, (iii) determinar se uma variante alélica tem ou não efeito na eficiência do *splicing*, (iv) efectuar a análise funcional de supostos elementos reguladores do processo de *splicing* previamente identificados por análise bioinformática, (v) identificar elementos necessários a uma regulação celular específica, (vi) identificar elementos requeridos na regulação do *splicing* por factores *trans-acting* específicos, e (vii) testar se uma mutação geneticamente ligada a uma doença afecta a eficiência do *splicing* de um exão adjacente ou causa activação de um local críptico de *splicing* (Cooper 2005; Singh & Cooper, 2006).

1.1.4.1 Construção de minigenes

A construção de minigenes é normalmente efectuada em vectores de expressão. Estes podem ser uma qualquer porção de DNA capaz de incorporar um fragmento de DNA de outro organismo, sendo depois este conjunto normalmente transferido para uma célula de um outro organismo. Existem vários tipos de vectores, contudo os mais utilizados são os vectores plasmídicos capazes de clonar fragmentos de DNA com um tamanho médio de 4 Kb. Um plasmídeo é uma molécula de DNA de cadeia dupla, pequena e circular, geralmente encontrado em algumas bactérias, leveduras, células de plantas ou outros tipos de organismos (Lewis 2001).

Nos estudos de *splicing*, os minigenes contêm habitualmente o fragmento genómico de interesse que inclui a região onde ocorre *splicing* alternativo, assim como as regiões genómicas flanqueantes. Este fragmento pode ser obtido através de amplificação por PCR directamente a partir do gDNA ou de um fragmento de DNA já clonado. O comprimento das regiões genómicas flanqueantes depende dos objectivos de cada análise em particular e pode variar entre as dezenas e centenas de nucleótidos. Sabe-se actualmente que os elementos básicos que regulam o processo de *splicing* encontram-se tipicamente entre 200-300 nucleótidos a jusante ou a montante do exão sob regulação. No entanto, existem casos em que estes elementos reguladores específicos foram encontrados a mais de um kilobase do *splice-site* regulado (Savkur *et al.*, 2001; Dredge & Darnell, 2003). Apesar deste facto, segundo Cooper (2005) um fragmento genómico que contenha até 300 nucleótidos de sequência intrónica a jusante e a montante do exão é um bom ponto de partida para efectuar a análise de elementos *cis-acting* (Figura 4). Porém, nos casos em que o fragmento inicial não permitiu a análise da regulação, de forma a observar a regulação do processo de *splicing* foi necessária a utilização de segmentos maiores de gDNA.

A amplificação do fragmento de DNA de interesse a ser clonado no minigene deverá ser efectuada utilizando oligonucleótidos desenhados de forma a conter no final da extremidade 5' os locais específicos para as enzimas de restrição que emparelham no locais de restrição existentes no plasmídeo seleccionado. Deste

modo, estas enzimas são normalmente escolhidas de acordo com a disponibilidade dos sítios de clonagem existentes no minigene (RS#1 e RS#2, Figura 4). Para além disso, será conveniente seleccionar duas enzimas de restrição que possuam a sua máxima eficácia de corte no mesmo tampão. Dado que muitas enzimas de restrição não cortam eficientemente nos locais que se encontram após o final do fragmento de interesse clonado, uma sequência de 6 nucleótidos (frequentemente ATATAT), deverá ser incluída para que estes sejam os primeiros nucleótidos do *primer*. Desta forma, o *primer* desenhado terá a partir da direcção 5' a sequência ATATAT, seguida pela sequência específica de restrição e pelos 18-21 nucleótidos complementares ao gene de interesse. Por outro lado, os locais de clonagem escolhidos e desenhados nos *primers* utilizados, devem permitir apenas a amplificação do produto de PCR de interesse e esta deve ser efectuada com uma DNA polimerase que funcione com elevada exactidão para evitar a introdução de substituições nucleotídicas.

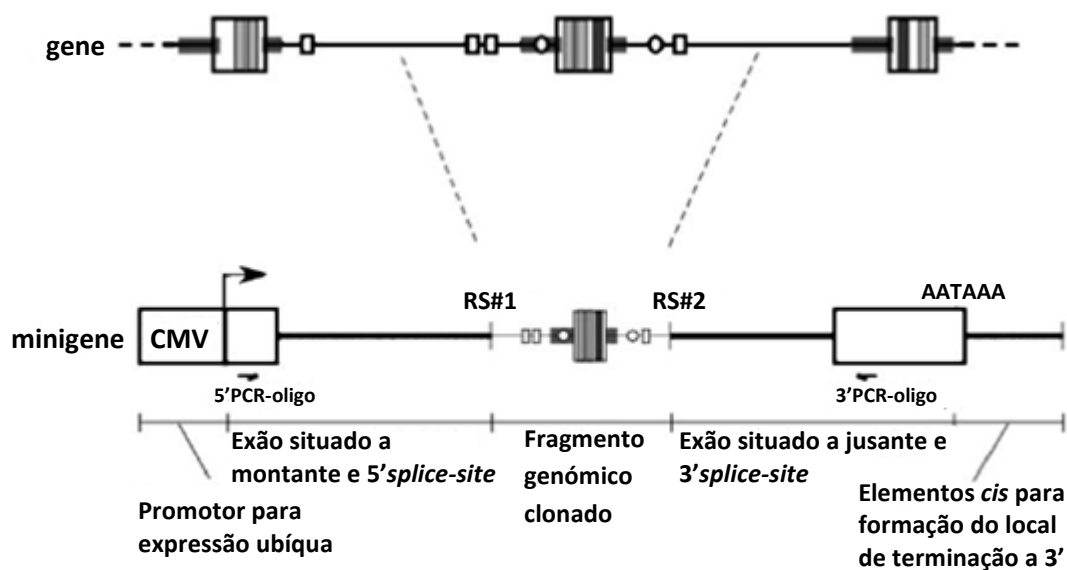


Figura 4: Esquema representativo da construção de um minigene num vector plasmídico. CMV – Citomegalovirus *transcriptional enhancer/promotor*; RS#1 e RS#2 – locais de restrição situados no plasmídeo. Adaptado de Cooper (2005).

1.1.4.2 Transfecção e obtenção de RNA

Para avaliar o funcionamento *in vivo* de um minigene construído com o fragmento de interesse, é necessário proceder-se à sua transfecção numa linha celular. A transfecção é o processo de introduzir DNA ou RNA exógenos em culturas de células eucarióticas, sendo uma técnica indispensável no estudo da função e expressão de proteínas, na modulação da expressão génica, na análise mutacional e na produção proteica. Existem vários métodos de transfecção, incluindo métodos físicos como a electroporação, microinjecção ou choque térmico e métodos bioquímicos que envolvem a utilização de fosfato de cálcio ($\text{Ca}_3(\text{PO}_4)_2$), polímeros, lipossomas ou dendrímeros (macromoléculas orgânicas altamente ramificadas) (Kendrem 1994). Um dos métodos mais eficientes e menos citotóxico para células epiteliais é o que utiliza lipossomas catiónicos como mediadores da transfecção. Os lipossomas são estruturas sintéticas de lípidos catiónicos com uma constituição similar às membranas celulares que têm à superfície carga positiva, o que lhes confere a capacidade de interagir de uma forma electrostática com o material genético carregado negativamente e de formar complexos (lipossoma/ácido nucleico) estáveis. Estes complexos são depois integrados no citoplasma da célula por endocitose, beneficiando da atracção inerente que, também, possuem com a membrana plasmática (Kendrem 1994; Almeida & Saraiva, 2001).

A transfecção pode ser transitória ou estável, embora para a maior parte das aplicações seja suficiente que o material genético seja expresso, apenas transitoriamente, isto é, sem integrar o genoma e desaparecendo após mitose celular (Kendrem 1994). Trata-se de um processo que pode ser efectuado em várias linhas celulares, sendo a sua escolha efectuada de acordo com os objectivos de cada estudo em particular. Geralmente, células com elevada eficiência de transfecção como HeLa, COS e HEK293 são utilizadas para definir as bases do *splicing* “não-constitutivo” (isto é, para identificar as características que resultam num processo de *splicing* ineficiente). A reacção de transfecção pode ser efectuada com diferentes reagentes disponíveis no mercado, porém, deverá ser escolhido um reagente que permita uma elevada eficiência de transfecção em múltiplas linhas celulares, fácil de usar e com nível de toxicidade baixo (Cooper 2005).

Para a identificação de elementos *cis-acting* necessários a uma regulação celular específica, a abordagem ideal passa por utilizar uma cultura celular que possa ser induzida a uma transição fisiológica associada com a transição que ocorre no processo de *splicing* sob investigação. Assim, idealmente, o minigene deverá ser capaz de demonstrar a mesma transição de *splicing* apresentada pelo gene endógeno (Cooper 2005).

1.1.5 Utilização de minigenes para o estudo de mutações de *splicing* – alguns exemplos

A técnica de clonagem e construção de minigenes é uma técnica bastante utilizada laboratorialmente para o estudo do processo de *splicing* em várias áreas da genética. Nos últimos anos, vários estudos têm sido realizados utilizando esta técnica com o objectivo de compreender as alterações provocadas por mutações de *splicing* ocorridas tanto em locais exónicos como intrónicos, como de contribuir para um aumento do conhecimento dos mecanismos que regulam o processo de *splicing* de um modo geral.

Existem vários trabalhos na literatura (Cartegni & Krainer, 2002; Kashima & Manley, 2003; Cartegni *et al.*, 2006) realizados para os genes *SMN1* (*Survival of motor neuron 1*) e *SMN2* (*Survival of motor neuron 2*) ambos associados à patologia neurodegenerativa – atrofia muscular espinhal (*Spinal muscular atrophy* - SMA) que é causada pela perda das cópias funcionais do gene *SMN1*. O gene *SMN2* é um gene parálogo idêntico ao *SMN1* e a sua sequência é igual em indivíduos normais e portadores da doença. Embora fosse esperado que os dois genes codificassem proteínas indistinguíveis, a proteína *SMN2* é apenas parcialmente funcional, sendo suficiente para uma actividade normal na maioria das células mas não nos neurónios motores. O gene *SMN2* é eficientemente transcrito, mas não permite compensar a perda da proteína *SMN1* devido a uma substituição nucleotídica C>T silenciosa que ocorre ao nível do exão 7 e que provoca alterações no *splicing* do pré-mRNA, causando o *skipping* deste exão e levando à síntese de uma proteína instável. Para tentar perceber a diferença observada no processo de *splicing* dos genes *SMN1* e *SMN2*, alguns estudos têm sido efectuados tendo sido propostos dois

modelos. Um deles propõe a existência de um ESE no exão 7 do gene *SMN1* que é reconhecido pela proteína SR – AFS/SF2 e que a substituição nucleotídica C>T que ocorre no gene *SMN2* interrompe esse ESE resultando no *skipping* do exão 7. Estudos de análise mutacional, transfecção e análise do *splicing in vivo* foram efectuados apoiando a hipótese de que a proteína ASF/SF2 exerce uma função importante na inclusão do exão 7 no gene *SMN1* que não é mantida no gene *SMN2*. Um outro modelo sugere que a presença do nucleótido T no exão 7 do gene *SMN2* origina a criação de um ESS que é reconhecido pela proteína hnRNP A1 levando à exclusão do exão 7. Esta hipótese foi testada pelo silenciamento (por RNA de interferência - RNAi) da proteína hnRNP A1 permitindo verificar um aumento na inclusão do exão 7, o que sugere que esta proteína poderá estar envolvida na inibição da inclusão do exão 7 no gene *SMN2*. Estes dois modelos não são necessariamente incompatíveis, uma vez que tanto a perda de um ESE específico para a proteína ASF/SF2, como a criação simultânea de um motivo de ligação para a proteína hnRNP A1, poderão em princípio contribuir para o *skipping* do exão 7 no gene *SMN2*. Na realidade, estas duas proteínas parecem antagonizar-se numa competição aparentemente baseada nas suas concentrações relativas e propriedades de ligação ao RNA.

Dado que a gravidade da patologia SMA é diferente de acordo com o défice de funcionalidade da proteína SMN, o aumento da inclusão do exão 7 para o gene *SMN2* poderá diminuir a gravidade desta doença através da produção da proteína *SMN2* funcional. Neste sentido, alguns trabalhos têm sido realizados (Hua *et al.*, 2008; Singh *et al.*, 2009), utilizando por exemplo Oligonucleótidos Antisense, com o objectivo terapêutico de reverter a exclusão do exão 7 *SMN2* e assim aumentar os níveis de expressão de uma proteína funcional. Estes estudos têm demonstrado tanto em linhas celulares, como em modelos animais (ratinho) a recuperação de uma proteína funcional, abrindo assim uma possibilidade terapêutica e de cura para esta doença.

Num outro estudo, Gonçalves *et al.* (2009) efectuaram a análise do impacto no processo de *splicing* de uma nova mutação *missense* (c.1918C>G) no exão 14 do gene supressor tumoral *APC* (*Adenomatous polyposis coli*), identificada num doente diagnosticado com *Familial adenomatous polyposis* (FAP). A análise *in vitro* do cDNA

do doente revelou o *skipping* completo do exão 14 nos transcritos mutantes dando origem a uma alteração na grelha de leitura e consequentemente a um codão STOP prematuro. Deste modo, para determinar se a mutação referida foi suficiente para causar a ausência do exão 14 foram construídos minigenes para a sequência *wild-type* e mutante abrangendo o exão 14 e parte dos intrões 13 e 14. A análise funcional *in vivo* para ambos os minigenes permitiu verificar a inserção do exão 14 numa porção significativa dos transcritos derivados do minigene *wild-type*, e o *skipping* total do exão 14 nos transcritos resultantes do minigene com a sequência mutante, demonstrando que a mutação *missense* em estudo foi suficiente para causar a ausência deste exão no transcrito produzido. Verificaram ainda que a mutação estudada interrompia um local para a ligação das proteínas – SRp55 e ASF/SF2 e criava um motivo de ligação para a proteína hnRNP A1. Para validar estas previsões os autores realizaram o silenciamento destas proteínas (RNAi) nas células onde foram transfectados os minigenes construídos, tendo verificado que apenas a proteína ASF/SF2 foi necessária para a inclusão do exão 14. Observaram ainda uma ligeira recuperação da inclusão deste exão por sobreexpressão da proteína ASF/SF2 em células transfectadas com o minigene mutante, apoiando a hipótese de que o motivo do ESE com ligação para a proteína ASF/SF2 é alterado na presença da mutação, podendo assim ser este o mecanismo pelo qual a mutação c.1918C>G altera o processo normal de *splicing* do pré-mRNA na região afectada do gene *APC*. Este trabalho demonstra assim a importância do uso de minigenes repórter como uma ferramenta útil para o estudo das alterações provocadas por uma determinada mutação no processo de *splicing*.

Num outro trabalho, onde foi também analisado o impacto de mutações no processo de *splicing* do gene *PMM2* (*Phosphomannomutase 2*) envolvido na Doença Congénita da Glicosilação tipo I (Vega *et al.*, 2009), os autores recorreram não só à construção de minigenes como base para a sua realização, como também a outros ensaios (análise mutacional, sobreexpressão de proteínas e análises de *splicing in vitro* e *in vivo*), com o objectivo, de estudar também o efeito de mutações específicas no mecanismo de *splicing*. Após a análise da mutação c.640-15479C>T verificaram que esta foi responsável pela activação de um pseudoexão com 123 pb a partir da sequência do intrão 7. Assim, com o objectivo de corrigir o defeito

causado por esta substituição nucleotídica, utilizaram oligonucleótidos morpholino antisense (AMOs) desenhados para os locais 5' e 3' crípticos utilizados na presença da mutação. Observaram experimentalmente que estes AMOs permitiram bloquear o acesso da maquinaria de *splicing* aos locais crípticos, originando a produção de um transcrito de mRNA correcto que foi eficientemente traduzido na proteína fosfomanomutase funcional.

No caso de outras patologias designadas colectivamente como distrofinopatias (Distrofia muscular de Duchenne (DMD); Distrofia muscular de Becker (BMD) e cardiomiopatia dilatada) têm também sido efectuados estudos de mutações de *splicing*. Estas patologias são causadas na maioria dos casos por deleções no gene da distrofina, embora um grande número de mutações pontuais (exónicas e intrónicas) sejam também apontadas como causa de doença por provocarem alterações no processo de *splicing*. Num estudo recente Disset e colaboradores (Disset *et al.*, 2006), analisaram uma mutação nonsense (c.4250T>A – p.Leu1417X) no gene da distrofina num doente com um fenótipo intermédio (BMD). Verificaram após análises *in vitro* e *in vivo* recorrendo ao uso de minigenes que esta substituição nucleotídica no exão 31 para além de criar um codão STOP, originou também a criação de um ESS com ligação para a proteína hnRNP A1 resultando no *skipping* do exão 31. Identificaram também a presença de um ESE com ligação à proteína Tra2 β situado a jusante do motivo de ligação para a proteína hnRNP A1. Os autores verificaram que a presença desse motivo ESE é necessária para a incorporação do exão 31 no transcrito final, tanto na sequência *wild-type* de um controlo como na sequência mutada do doente, permitindo concluir que a regulação da inserção deste exão no mRNA maduro dependerá da interacção entre diferentes factores *trans-acting*.

Num outro estudo realizado por Alter *et al.* (2006) foi verificado que uma deleção envolvendo três exões (48 a 50) no gene da distrofina conduziu ao aparecimento de um codão STOP ao nível do exão 51 devido a uma alteração na grelha de leitura. Com o objectivo de reverter esta alteração os autores recorreram ao uso de oligonucleótidos antisense num modelo animal (ratinho) direccionando-os para o local de *splicing* 3' do exão 51 para prevenir a sua inclusão no transcrito, restabelecendo assim a grelha de leitura e possibilitando a codificação de uma

proteína parcialmente funcional que poderia assim diminuir a gravidade da doença. Para esta patologia num estudo recente realizado por van Deutekom *et al.* (2007), foi já administrado por via intramuscular um oligonucleótido antisense denominado PRO051 em 4 doentes que possuíam deleções no exão 50 provocando o aparecimento de um codão STOP prematuro no exão 51, com o objectivo de provocar o *skipping* deste exão e restaurar a grelha de leitura para que pudesse ser produzida uma proteína parcialmente funcional. Verificaram que em todos os doentes ocorreu o *skipping* do exão pretendido e que foi produzida a proteína distrofina (3% a 12% em extractos de proteína total) sugerindo que estas abordagens terapêuticas poderão ser viáveis.

Podemos assim concluir que estes trabalhos são extremamente importantes para a identificação de elementos *cis-acting* e factores *trans-acting* envolvidos não só na regulação do *splicing* em casos específicos onde ocorreram alterações devido à presença de mutações, como também para a compreensão do processo de *splicing* de um modo geral. O conhecimento adquirido destes estudos tem fornecido uma série de potenciais alvos terapêuticos para os quais têm sido desenvolvidas abordagens na tentativa de corrigir os erros provocados por determinada mutação de *splicing* (Cooper *et al.*, 2009).

Capítulo II

Objetivos

2. Objectivos

O presente trabalho teve como principal objectivo a análise da regulação dos mecanismos de *splicing* no gene *IDS*, através do estudo do impacto da mutação c.257C>T no exão 3 deste gene.

Mais especificamente os objectivos deste estudo foram:

- 1) Utilização de ferramentas bioinformáticas para a identificação dos elementos reguladores do *splicing* alterados na presença da mutação c.257C>T;
- 2) Análise funcional do processo de *splicing* no gene *IDS* na presença desta mutação, através da construção de minigenes repórter recorrendo ao uso de vectores plasmídicos e transfecção de uma linha celular de cancro colorectal DT-1;
- 3) Identificação dos elementos reguladores do *splicing* alterados pela mutação c.257C>T, através de estudos de sobreexpressão e silenciamento (RNAi) das proteínas (SR e hnRNP) previstas bioinformaticamente como implicadas na regulação do *splicing* desta região do gene *IDS*.

De um modo global este trabalho teve também como propósito a implementação destas metodologias no grupo de investigação em Doenças Lisossomais de Sobrecarga, para sua futura utilização em outros trabalhos de investigação.

Capítulo III

Material e Métodos

3. Material e Métodos

3.1 Material biológico

O presente trabalho insere-se num estudo de caracterização molecular do gene *IDS* em doentes com diagnóstico bioquímico de Mucopolissacaridose tipo II (Síndrome de *Hunter*). O diagnóstico desta patologia foi efectuado na Unidade de Bioquímica - Genética do Centro de Genética Médica Doutor Jacinto de Magalhães – INSA, I.P, através da determinação da actividade enzimática da Iduronato-2-sulfatase (IDS) em fibroblastos.

A amostra de um doente com uma mutação no exão 3 do gene *IDS* (c.257C>T; p.P86L) foi analisada e como material biológico foi utilizada uma linha celular de fibroblastos de pele cultivados e crioconservados em azoto líquido. Como controlo, foi utilizada uma linha celular de fibroblastos de um indivíduo saudável.

3.2 Metodologia experimental

3.2.1 Cultura celular de fibroblastos e obtenção de *pellets* celulares

Para a realização deste trabalho, as linhas celulares de fibroblastos pertencentes ao doente em estudo e ao indivíduo controlo foram repostas em cultura de modo a obter um número de células suficiente para a extracção de DNA e RNA.

Todo o procedimento foi realizado em câmara de fluxo laminar, sendo o material não esterilizado devidamente desinfectado, antes da entrada na câmara, com álcool a 70% para evitar contaminações.

Para reposição em cultura, as células crioconservadas foram ressuspensas no meio *Dulbecco's Modified Eagle Medium* (DMEM) [Gibco Invitrogen, Carlsbad, CA, USA], complementado com 10% de *Fetal Bovine Serum* (FBS) [Gibco®] e antibióticos (10 000 U/mL de estreptomicina e penicilina, 1 mg/mL de fungizona), a 37°C, numa atmosfera humidificada com 5% de CO₂. Até à obtenção de uma cultura com

número suficiente de células para extracção de DNA ou RNA, procedeu-se regularmente à renovação do meio de cultura.

Quando a cultura celular de fibroblastos se apresentava confluenta e uniformemente distribuída, o meio de cultura foi retirado e as monocamadas celulares lavadas com solução salina de tampão fosfato (PBS), seguindo-se a adição de tripsina [Gibco®] de modo a dissociar as células da caixa de cultura.

Adicionou-se posteriormente meio de cultura à suspensão celular obtida e procedeu-se à sua transferência para um tubo de centrífuga. A suspensão foi, então, lavada por duas vezes, centrifugada a 2000g e os depósitos celulares obtidos foram armazenados a -80°C, até se proceder à extracção de DNA genómico (gDNA) ou de RNA total.

3.2.2 Extracção de DNA genómico

A extracção de gDNA foi efectuada por extracção automática no aparelho *Bio Robot EZ1* [Qiagen], utilizando um *kit* comercial de extracção específico *EZ1 DNA tissue* [Qiagen]. Na utilização deste *kit* foi seguido o protocolo indicado pelo fabricante.

A determinação da concentração (ng/μL) e da pureza do DNA total foi obtida por quantificação no espectrofotómetro UV-Vis *NanoDrop® ND-1000*. Além da determinação da concentração de DNA, foi igualmente avaliado o grau de contaminação proteica através do valor da razão de absorvância – Abs 260nm/280nm. Os valores localizados no intervalo 1,80-2,20 foram considerados aceitáveis. Após estes procedimentos, o DNA foi conservado a -20°C até à sua utilização.

3.2.3 Extracção de RNA total

Para a obtenção de RNA a partir das amostras biológicas mencionadas anteriormente foi utilizado o *kit High Pure RNA Isolation* [Roche], tendo-se procedido resumidamente do seguinte modo: **1)** extracção do conteúdo celular utilizando o

tampão de lise (*Lysis Binding Buffer*), que permite ao mesmo tempo a inactivação de RNases; **2)** separação selectiva dos ácidos nucleicos por ligação às fibras de vidro dos tubos fornecidos (*High Pure Filter Tube*); **3)** eliminação das moléculas de DNA contaminantes por adição de uma solução contendo a enzima DNase I; **4)** de forma a eliminar todas as substâncias que possam interferir com a reacção de RT-PCR, os ácidos nucleicos que se encontrassem ligados às fibras de vidro do tubo foram eluídos utilizando o tampão *Inhibitor Removal Buffer*; **5)** lavagem para eliminação de ácidos nucleicos que ainda se encontrem ligados, assim como para purificação de sais, proteínas e outras impurezas celulares; **6)** recuperação do RNA purificado por eluição com o tampão *Elution Buffer*; e por fim **7)** armazenamento do RNA extraído por conservação a -80°C até à sua utilização.

A determinação da concentração (ng/μL) e da pureza do RNA total foi também obtida por quantificação no espectrofotómetro UV-Vis *NanoDrop® ND-1000*.

3.2.4 Análise mutacional do cDNA do gene *IDS*

Procedeu-se à análise do cDNA *IDS* - no doente em estudo e na amostra controlo. A partir de cada uma das amostras de RNA total, foi sintetizado cDNA (como descrito em 3.2.4.1) e utilizados diferentes oligonucleótidos para a amplificação específica do cDNA *IDS*. Os produtos obtidos foram posteriormente analisados e sequenciados como descrito, respectivamente, em 3.2.4.3 e 3.2.4.4, tendo sido confirmada a presença da mutação (c.257C>T) no cDNA do doente. A presença desta mutação foi também confirmada no gDNA do doente.

3.2.4.1 Síntese de cDNA

A cadeia de DNA complementar ao mRNA foi sintetizada a partir do RNA total, utilizando o *kit First-Strand cDNA Synthesis* [Amersham Biosciences]. A reacção de síntese é catalisada pela enzima retrovítica transcriptase reversa, incorporada na mistura do *kit* utilizado.

O procedimento de síntese obedeceu à seguinte sequência:

1) Destruição das estruturas secundárias do RNA, por aquecimento da solução contendo RNA (com uma concentração final de 1 µg) a 60°C durante 10 minutos.

2) Adição da mistura de reacção (*bulk first-strand reaction mix*, primer Not I-(dT)₁₈ diluído e uma solução de DTT) e incubação a 37°C durante 1 hora, seguida de nova incubação a 93°C durante 10 minutos.

3.2.4.2 Amplificação por PCR dos fragmentos de cDNA *IDS*

O cDNA do gene *IDS* foi amplificado em 3 fragmentos tal como descrito em Alves *et al.*, 2006 (Tabela I). Para todas as reacções de amplificação foi utilizada a mistura comercial *ImmoMix Red 1x* [Bioline] (contendo DNA polimerase IMMOLASE™, (NH₄)₂SO₄, Tween 20, MgCl₂, dNTPs e tampão Tris-HCl) e *primers* (*Forward* e *Reverse*) a uma concentração de 0,5 µM. A utilização desta mistura de amplificação apresentou diversas vantagens, na medida em que diminuiu consideravelmente o número de manipulações efectuadas, bem como os riscos de contaminação. Na mistura de reacção utilizaram-se 5 µL de cDNA para um volume final de 50 µL. O controlo negativo (branco do ensaio) foi realizado através da substituição da amostra de cDNA por água desionizada esterilizada.

O programa de amplificação utilizado consistiu num ciclo inicial de 7 minutos a 95°C, seguido de 30 ciclos de 45 segundos a 94°C, 45 segundos a 60°C, e 2 minutos a 72°C, finalizando-se a reacção com um ciclo de extensão de 10 minutos a 72°C.

Tabela I. Sequência dos *primers* utilizados na amplificação dos fragmentos cDNA *IDS*.

Fragmento de cDNA <i>IDS</i>	Sequência dos <i>primers</i> (5'→3')	Tamanho do produto amplificado
Fragmento 1 (exões 1 a 5)	1F: 5'-CTGTGTTGCGCAGTCTTCAT-3' 1R: 5'-GGGGTATCTGAAGGGGATGT-3'	829 pb
Fragmento 2 (exões 5 a 8)	2F: 5'-CTGTGGATGTGCTGGATGTT-3' 2R: 5'-GGGTCGAGGTAAGGGAAAAG-3'	595 pb
Fragmento 3 (exões 8 a 9)	3F: 5'-GATGTTGCTACCCATGTTCC-3' 3R: 5'-CAAAACGACCAGCTCTAACTC-3'	698 pb

pb – pares de bases; F – *forward*; R – *reverse*

3.2.4.3 Análise da reacção de PCR e purificação dos produtos amplificados

O sucesso da amplificação dos diversos fragmentos pretendidos foi verificado através da separação destes por electroforese em gel de agarose a 2%, em tampão TAE 1x, juntamente com um marcador de pesos moleculares de 100 pb [Fermentas].

A visualização dos fragmentos pretendidos foi efectuada num transiluminador de luz ultra-violeta e sempre que foram amplificados apenas os fragmentos pretendidos, os produtos de PCR foram purificados em colunas *Illustra MicroSpin S300Hr* [GE Healthcare] de acordo com o protocolo fornecido pelo fabricante. Nos casos em que ocorreram fragmentos de amplificação inespecíficos ou mais de uma banda de interesse, a banda correspondente ao fragmento pretendido foi excisada e purificada com o *kit Perfectprep Gel Cleanup* [Eppendorff], seguindo as indicações do fabricante.

3.2.4.4 Sequenciação do cDNA *IDS*

A preparação das amostras para sequenciação automática iniciou-se pela reacção de sequenciação (PCR assimétrico). Para tal, utilizou-se uma mistura comercial *BigDye® Terminator Cycle Sequencing kit 2.0* [ABI PRISM, PE Applied

Biosystems] que contém dNTPs, ddNTPs marcados com 4 fluorocromos distintos, tampão e uma DNA polimerase, à qual se adiciona o produto de PCR previamente purificado e um dos *primers* (Tabela II).

Tabela II. Mistura de reacção utilizada no PCR de sequenciação.

	Volume	Concentração
Mix de Sequenciação (<i>Big Dye</i>®)	2,0 µL	-
<i>Primer Forward</i> ou <i>Primer Reverse</i>	0,5 µL	3,2 pmol/µL
DNA purificado	2,5 - 3 µL	-

Nota: A quantidade de amostra utilizada variou consoante a intensidade da banda visualizada em gel após a purificação dos produtos.

A reacção de PCR assimétrico foi efectuada utilizando as condições indicadas na Tabela III.

Tabela III. Condições do PCR de sequenciação.

Temperatura	Tempo	Ciclos
96 °C	3 min	1
96 °C	10 seg	25
50 °C	5 seg	
60 °C	4 min	

Nota: Nos casos em que a amplificação de um determinado fragmento foi menos eficiente, o número de ciclos foi aumentado para 35.

Os fragmentos amplificados foram purificados em colunas preenchidas com resina Sephadex 7% [GE Healthcare], através de centrifugação a 4400 rpm durante 4 minutos. Seguidamente, foi adicionada a cada amostra de produto purificado 15 µL de Hi-Di Formamida Desionizada [Applied Biosystems].

A análise dos fragmentos foi realizada segundo o método desenvolvido por Sanger *et al.*, 1977, num sequenciador automático *ABI PRISM™ 3130XL Genetic Analyser* de 16 capilares de 36 cm [Applied Biosystems].

Na análise das sequências nucleotídicas obtidas foram utilizadas como referência as sequências: ENSG00000010404 (sequência genómica do gene *IDS*) e

ENST00000340855 (sequência do cDNA *IDS*), ambas disponíveis na base de dados *Ensembl* (<http://www.ensembl.org>). O alinhamento de cada sequência obtida com a sequência de referência foi efectuado com o *software ClustalW* (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>).

3.2.5 Análise Bioinformática

A previsão dos padrões de *splicing* para o gene *IDS* na presença da mutação c.257C>T foi efectuada com o *software* GENSCAN (<http://genes.mit.edu/GENSCAN.html>) (Burge & Karlin, 1997). Os valores de *score* para os locais de *splicing* foram calculados utilizando o programa *MaxEntScan* (http://www.genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html) (Yeo & Burge, 2004). Segundo este programa, os valores ideais de *score* para os locais de *splicing* são: 11,81 (5'ss) e 13,59 (3'ss).

Para efectuar a previsão de alterações em ESEs foi utilizado o programa *ESE finder 3.0* (http://rulai.cshl.edu/cgi-bin/tools/ESE3/ese_finder.cgi?process=home) (Cartegni *et al.*, 2003). A identificação de possíveis locais de ligação de factores de *splicing* como proteínas SR e hnRNP foi realizada com a ferramenta bioinformática *Splicing Rainbow* (<http://www.ebi.ac.uk/asd-srv/wb.cgi?method=8>) (Stamm *et al.*, 2006).

3.2.6 Análise funcional do impacto da mutação c.257C>T no processo de *splicing* do gene *IDS* – Clonagem e construção de minigenes repórter

Com o objectivo de realizar a análise funcional do processo de *splicing* no gene *IDS*, na presença da mutação c.257C>T, foi efectuada a construção de minigenes repórter recorrendo ao uso de vectores plasmídicos.

3.2.6.1 Amplificação por PCR dos fragmentos de DNA genómico de interesse

Com o objectivo de estudar a mutação de *splicing* referida, que se localiza no exão 3 do gene *IDS*, procedeu-se à amplificação por PCR dos fragmentos de gDNA de interesse tanto do doente como do controlo.

Uma vez que existe um pseudogene *IDS* que possui uma elevada homologia com o gene *IDS* na zona correspondente aos exões 2 e 3 e ao intrão 2 os *primers* específicos para amplificação do fragmento de interesse foram desenhados entre o intrão 1 e o intrão 3 para que o fragmento amplificado fosse o correspondente ao gene e não ao pseudogene. Para além disso, dado que cada fragmento amplificado foi clonado num vector plasmídico com locais de corte específicos para as enzimas de restrição *Bam*HI e *Xho*I, as “caudas” com a sequência consenso da enzima *Bam*HI e *Xho*I foram adicionadas, respectivamente ao *primer forward* e *reverse* para que cada fragmento amplificado contivesse a sequência de restrição específica (Figura 5 e Tabela IV). Os programas *primer3* e *oligonucleotide properties calculator* (<http://frodo.wi.mit.edu/primer3/> e <http://www.basic.northwestern.edu/biotools/oligocalc.html> respectivamente) foram utilizados para analisar as características dos *primers* desenhados, relativamente a diversos parâmetros tais como dimerização, qualidade e temperatura de *annealing*.

Nas reacções de amplificação foi utilizada a mistura comercial *Immomix Red 1x* [Bioline], *primers forward* e *reverse* com uma concentração de 1 μ M e aproximadamente 40 ng de gDNA para um volume final de 25 μ L (Tabela IV).

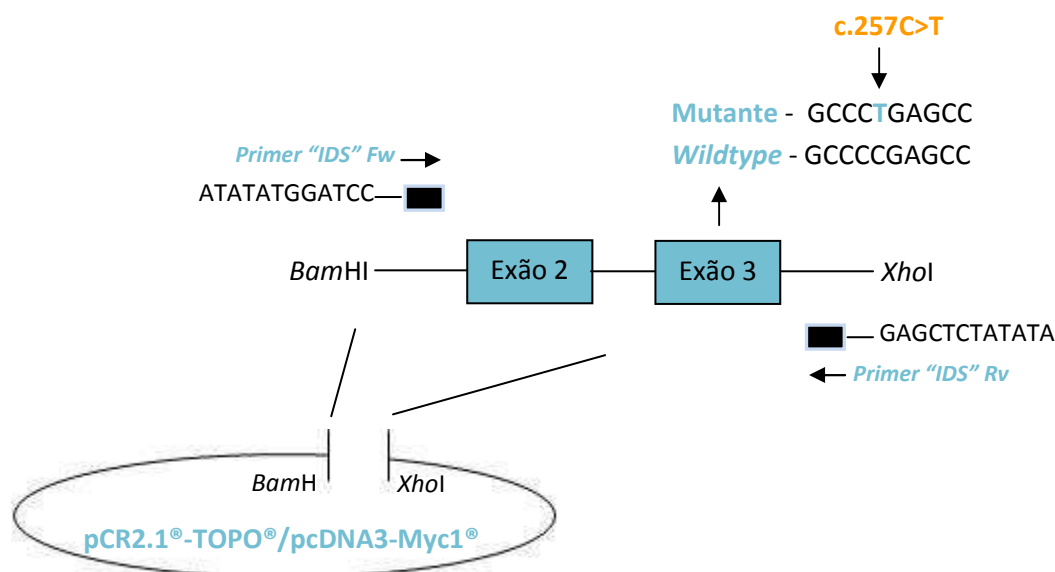


Figura 5: Esquema representativo dos locais de emparelhamento dos *primers* desenhados para amplificação dos fragmentos mutante e *wildtype* do gene *IDS*, assim como os locais de inserção nos vectores pCR2.1®-TOPO® e pcDNA3-Myc1®.

Tabela IV. Sequência dos *primers* utilizados na amplificação dos fragmentos de gDNA *IDS* (normal e com a mutação c. 257C>T) e respectivas condições de amplificação por PCR.

Sequência dos <i>primers</i> "IDS" (5'→3')	Tamanho	Condições de PCR
	do produto amplificado	
F - ATATATGGATCCTCCAGCCTTGGGCCTCTT R - ATATATCTCGAGGAATGCTGGATTGAGACA	1137 pb	95°C 7 min 94°C 45 seg △57°-63°C 45 seg 72°C 1 min 72°C 7 min

Abreviaturas: F – refere-se à sequência nucleotídica da cadeia orientada de 5' → 3' e R – à sequência nucleotídica da cadeia complementar; pb – pares de bases

3.2.6.2 Análise da reacção de PCR e Sequenciação

A amplificação dos fragmentos pretendidos foi verificada através da sua separação por electroforese em gel de agarose 1%, em tampão TAE 1x, juntamente com um marcador de pesos moleculares de 1Kb [Fermentas]. Após visualização dos fragmentos em transiluminador ultravioleta, os produtos foram purificados e

sequeenciados conforme o procedimento anteriormente descrito para o cDNA *IDS*, respectivamente em 3.2.4.3 e 3.2.4.4.

3.2.6.3 Construção de minigenes repórter intermédios

Com o objectivo de elevar o número de cópias dos fragmentos de interesse a clonar nos minigenes finais, procedeu-se à construção de minigenes intermédios através da sub-clonagem dos fragmentos amplificados de gDNA do doente e do controlo no vector plasmídico pCR2.1®-TOPO® [Invitrogen™] (Figura 6).

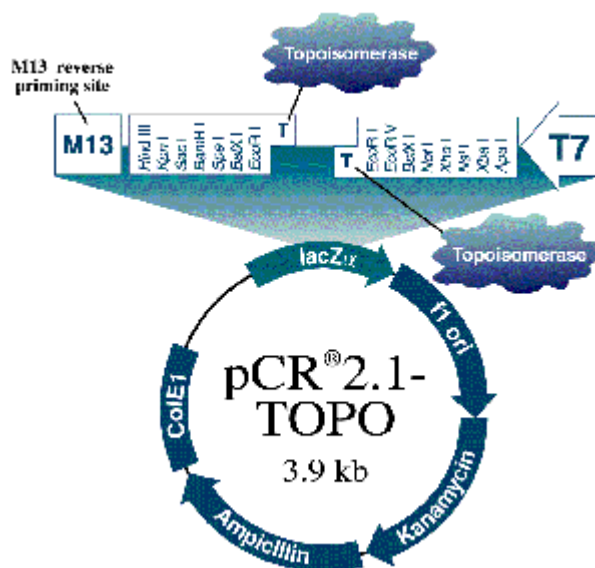


Figura 6: Vector plasmídico pCR2.1®-TOPO®

3.2.6.3.1 Sub-clonagem dos fragmentos de interesse obtidos por PCR no vector pCR2.1®-TOPO®

Após a reacção de amplificação, os produtos de PCR frescos foram clonados no vector pCR2.1®-TOPO® [Invitrogen™] de acordo com a reacção apresentada na Tabela seguinte:

Tabela V. Condições utilizadas na clonagem dos fragmentos de interesse no vector pCR2.1®-TOPO® [Invitrogen™].

Reagente	Quantidade
Produtos de PCR frescos	4 µL
Solução salina	1 µL
Água esterilizada	1 µL
Vector TOPO®	1 µL
Volume final	7 µL

Nota: O volume de produto de PCR a adicionar pode variar (0,5-4µL) de acordo com a intensidade das bandas visualizadas em gel de agarose. Neste caso foi utilizado o volume máximo possível: 4 µL.

Depois de constituída a mistura de cada um dos casos, foi agitada suavemente e incubada durante 5 minutos à temperatura ambiente. Após esse tempo foi colocada em gelo até se proceder à reacção de transformação em bactérias quimicamente competentes.

3.2.6.3.2 Transformação de células quimicamente competentes

As reacções de clonagem do fragmento *IDS* do doente e do controlo foram utilizadas para transformar bactérias quimicamente competentes da estirpe TOP10 de *Escherichia coli* (*One Shot® Chemically Competent Cells*) [Invitrogen™].

O processo de transformação foi efectuado de acordo com o procedimento seguinte: a cada alíquota de 50 µL de bactérias competentes adicionaram-se 5 µL de reacção de clonagem, seguindo-se um período de incubação de 30 minutos em gelo. Cada mistura foi, depois, sujeita a um choque térmico durante 45 segundos a 42°C, sendo colocada de imediato no gelo por 2 minutos. Após este período adicionaram-se, à chama, 250 µL (5x o volume de bactérias) de meio SOC (*Super Optimal broth with Catabolite repression*) pré-aquecido à temperatura ambiente e, de seguida, cada mistura foi incubada durante 1 hora a 37°C, com agitação constante (~220 rpm num agitador orbital). Posteriormente, 50 a 200 µL de cada reacção de transformação foram plaqueados em placas, pré-aquecidas a 37°C,

contendo meio LB-Agar (*Luria Bertani*) com ampicilina (Amp, 50 µg/mL) [Sigma], assim como 40 µL de substrato cromogénico (X-gal, 50 mg/mL) [Promega] para a enzima β-galactosidase. Por fim as placas foram colocadas a 37°C, durante um período de 16 a 18 horas. A figura 7 ilustra o processo de transformação descrito.

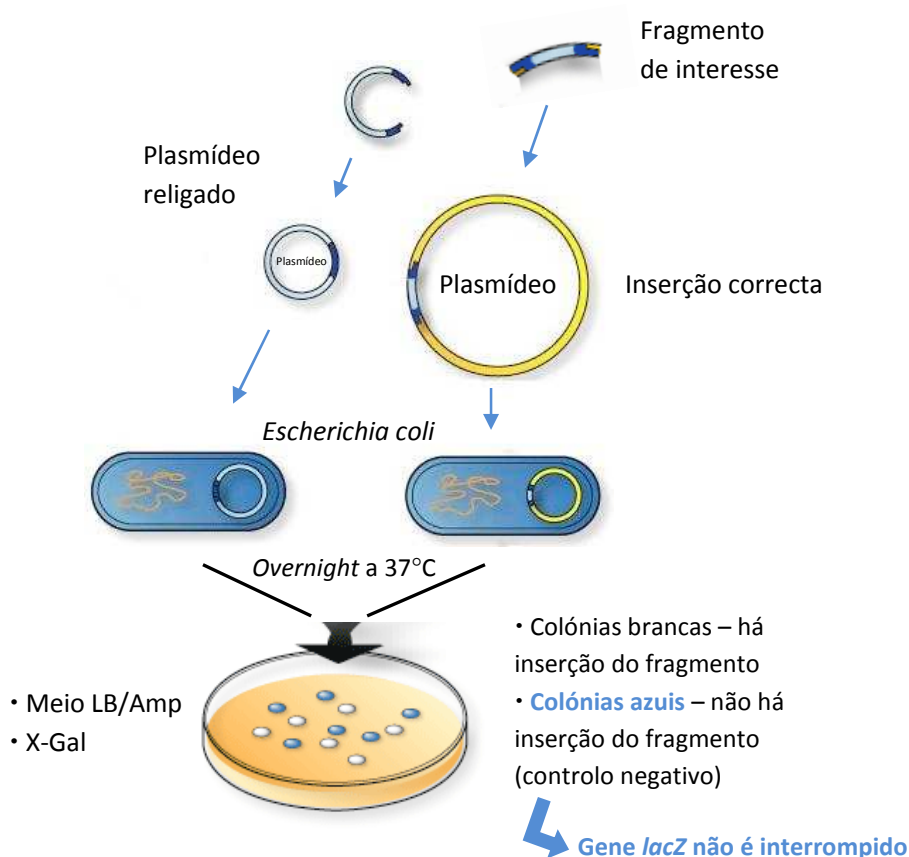


Figura 7: Esquema ilustrativo do processo de transformação do plasmídeo pCR2.1®-TOPO® em bactérias quimicamente competentes de *Escherichia coli*.

3.2.6.3.3 Obtenção de minigenes intermédios construídos no vector pCR2.1®-TOPO®

- Selecção de colónias por PCR

Após o tempo de incubação, cada placa foi analisada quanto à presença de colónias positivas (i.e., colónias brancas que integraram plasmídeos com os fragmentos de interesse) e colónias negativas (i.e., colónias azuis que integraram plasmídeos sem os fragmentos de interesse) por PCR. Para efectuar esta selecção, procedeu-se à diluição

de um número médio de 15 colónias brancas e 1 colónia azul em 15 µL de água desionizada esterilizada por cada tubo *ependorf*. Para todas as reacções de amplificação foi utilizada a mistura comercial *ImmoMix Red 1x* [Bioline], os *primers* T7 *forward* e M13 *reverse* a uma concentração de 0,25 µM e 5 µL de cada amostra contendo DNA plasmídico diluído para um volume final de 25 µL.

O programa de amplificação utilizado consistiu num ciclo inicial de 10 minutos a 95°C, seguido de 35 ciclos de 45 segundos a 94°C, 45 segundos a 54°C, e 1 minuto e 30 segundos a 72°C, finalizando-se a reacção com um ciclo de extensão de 15 minutos a 72°C.

A amplificação dos fragmentos pretendidos foi verificada através da sua separação por electroforese em gel de agarose a 1%, em tampão TAE 1x, juntamente com um marcador de pesos moleculares de 1Kb [Fermentas].

- Cultura de transformantes

Após a selecção por PCR, sempre que se obtiveram colónias positivas, procedeu-se à inoculação de 5 µL de cada amostra em tubos de Falcon contendo 5 mL de meio LB líquido e 5 µL do antibiótico de resistência do plasmídeo utilizado (Ampicilina, 50 µg/mL). Os tubos foram depois colocados, em cultura, a 37°C com agitação constante (220rpm), cerca de 16 horas.

- Isolamento, purificação e quantificação de transformantes

O DNA plasmídico de todas as culturas líquidas foi extraído das células bacterianas com o *kit peqGOLD Plasmid Miniprep Kit I* [peqLab, Biotechnologie GmbH] seguindo as instruções do fabricante. O procedimento realizado com este *kit* baseia-se num método modificado de lise alcalina (SDS) das células bacterianas que permite separar selectivamente o DNA cromossómico, proteínas celulares e outros contaminantes do DNA plasmídico. Este é depois purificado por adsorção a uma matriz de sílica (HiBind® matrix) e recuperado por eluição em água desionizada esterilizada ou tampão TE.

Depois de constituída cada *miniprep* com o DNA plasmídico extraído, procedeu-se à determinação das suas concentrações (ng/µL) e pureza por

quantificação no espectrofotómetro UV-Vis *NanoDrop*® *ND-1000*. Após estes procedimentos, o DNA foi conservado a -20°C até à sua utilização.

- Sequenciação de transformantes

Para verificar o sucesso do processo de clonagem foi efectuada sequenciação dos minigenes intermédios segundo o procedimento descrito em 3.2.4.4. É de salientar que neste caso específico foram utilizados 200 ng de DNA plasmídico, os *primers* T7 *forward* e M13 *reverse* e que o programa de PCR sofreu uma ligeira alteração na temperatura e tempo de desnaturação inicial (98°C durante 10 minutos), tendo-se mantido todas as restantes condições mencionadas em 3.2.4.4.

3.2.6.4 Construção de minigenes repórter finais

Para a construção dos minigenes finais pretendidos foi utilizado o vector plasmídico pcDNA3-Myc1® [Invitrogen™] (Figura 8). Este vector foi escolhido, uma vez que possui na sua sequência os locais de corte específicos para as enzimas de restrição *Bam*HI [NEB®_{Inc}] e *Xho*I [NEB®_{Inc}]. Estes locais permitirão a inserção de cada um dos fragmentos de gDNA de interesse que contêm, também, na sua sequência os locais de restrição específicos adicionados pelo respectivo *primer* na reacção de amplificação inicial por PCR. O referido vector foi ainda escolhido por possuir na sua sequência codificação para o epítipo Myc. Este é importante pois permite efectuar a distinção do fragmento de interesse em estudo do existente no DNA endógeno das células aquando da sua amplificação por PCR após transfecção.

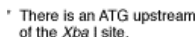


Figura 8: Vector plasmídico pcDNA3-Myc1®.

Os minigenes finais foram construídos de acordo com os procedimentos que a seguir se descrevem.

- Hidrólise enzimática dos vectores

Para a construção do minigene final com o fragmento de gDNA do doente foi efectuada a hidrólise enzimática de 1 µg de DNA do vector pcDNA3-Myc1® vazio, no qual se pretendia integrar o fragmento de interesse e de 3 µg de DNA do vector pCR2.1®-TOPO® que continha inserido individualmente o fragmento de gDNA de interesse do doente. A cada reacção de hidrólise foi adicionada a quantidade de DNA referida, 10 a 20 U da enzima *Bam*HI [NEB®_{Inc}] e *Xho*I [NEB®_{Inc}], um volume variável de tampão 10x [NEB®_{Inc}] adequado à dupla digestão das enzimas utilizadas, BSA 10x [NEB®_{Inc}] e água desionizada esterilizada para perfazer um volume final de 20 µL. As reacções foram depois incubadas durante cerca de 1 hora a 37°C (Figura 9).

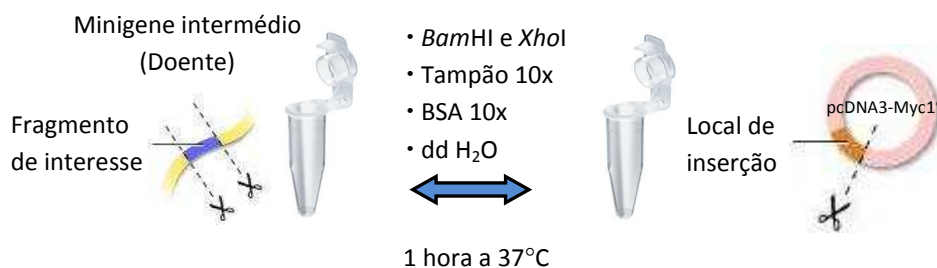


Figura 9: Esquema representativo da reacção de hidrólise do fragmento do gene *IDS* (doente) inserido no vector pCR2.1®-TOPO® e do vector pcDNA3-Myc1®.

Todas as reacções foram verificadas por electroforese em gel de agarose 1%, em tampão TAE 1x, junto com um marcador de pesos moleculares de 1Kb [Fermentas].

Quando concluída a digestão relativa ao fragmento de gDNA de interesse do doente, foi separada em gel de agarose *low melting* [Lonza] 1%, nas condições já referidas anteriormente, e as bandas pretendidas isoladas por excisão e purificadas com o kit *Perfectprep Gel Cleanup* [Eppendorff], segundo as indicações do fabricante. A reacção correspondente à hidrólise do vector vazio foi purificada directamente em colunas *Illustra MicroSpin S300Hr* [GE Healthcare] de acordo com o protocolo fornecido pelo fabricante.

- Ligação de produtos de hidrólise enzimática purificados

Após a obtenção dos produtos purificados das reacções de hidrólise enzimática, procedeu-se à sua ligação com recurso ao kit *Rapid DNA Ligation* [Roche]. No caso específico deste trabalho, a enzima T4 DNA ligase contida no kit foi utilizada para, na presença de ATP, efectuar a ligação entre o fragmento de interesse e o vector pcDNA3-Myc1® previamente digeridos.

Segundo o procedimento recomendado pelo fabricante, as reacções de ligação foram executadas em tubos *ependorf* onde se adicionaram 10 µL de tampão 1 e 2 µL de tampão 2, ambos incluídos no kit, 7 µL de DNA do fragmento de interesse, 1 µL de DNA do vector pcDNA3-Myc1® e 1 µL de enzima (5 U) para um volume final de 21 µL. Após homogeneização, as misturas foram incubadas durante 20 minutos à temperatura ambiente.

Os produtos de ligação obtidos foram depois transformados em bactérias quimicamente competentes da estirpe TOP10 de *Escherichia coli* (*One Shot® Chemically Competent Cells*) [Invitrogen™], segundo o procedimento descrito em 3.2.6.3.2. Contudo, neste caso o reagente X-gal [Promega] não foi utilizado pois o vector pcDNA3-Myc1® não possui o gene *lacZ*, logo não codifica a enzima β -galactosidase não sendo portanto este substrato necessário no meio (Figura 10).

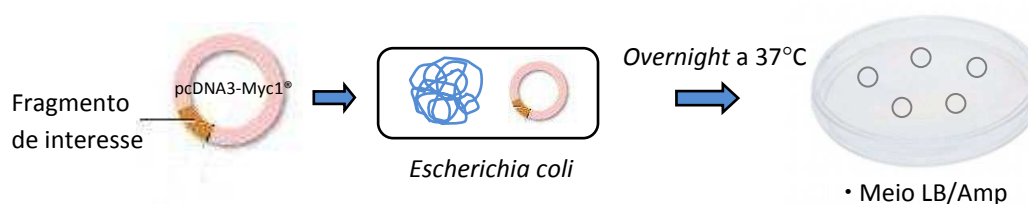


Figura 10: Esquema ilustrativo do processo de transformação do plasmídeo pcDNA3-Myc1® em bactérias quimicamente competentes de *Escherichia coli*.

Após a transformação foi efectuada a selecção de colónias positivas (i.e., colónias que integraram plasmídeos com o fragmento de interesse) por PCR com os *primers* pCMV5 *forward* (5'-GGGACTTCCAAAATGTCGTA-3') e pGHR1 *reverse* (5'-CCCACTGTCCTTCCTAATAAA-3') utilizando uma temperatura de *annealing* de 58°C. Todo o procedimento e restantes condições de PCR foram seguidas como mencionado para os minigenes intermédios no ponto 3.2.6.3.3.

Depois da selecção, sempre que foram detectadas colónias positivas efectuaram-se inóculos das amostras diluídas correspondentes para cultura do DNA plasmídico, procedendo-se depois ao seu isolamento, purificação e quantificação tal como descrito anteriormente.

Por fim, o DNA (200 ng) do minigene final obtido para o caso do doente foi amplificado por PCR para sequenciação com o *primer* T7 *forward* e "IDS" *reverse* desenhado para amplificação inicial do fragmento de interesse. As condições de tempo e temperatura utilizadas mantiveram-se como as já indicadas para o PCR de sequenciação dos minigenes intermédios.

Na tabela seguinte encontram-se sintetizadas as diversas etapas efectuadas para a clonagem dos fragmentos de gDNA do controlo e do doente nos plasmídeos pCR2.1®-TOPO® e pcDNA3-Myc1®.

Tabela VI. Etapas seguidas no processo de clonagem nos plasmídeos pCR2.1®-TOPO® e pcDNA3-Myc1®.

		PLASMÍDEOS	
		pCR2.1®-TOPO®	pcDNA3-Myc1®
ETAPAS	Hidrólise Enzimática	<i>Bam</i> HI e <i>Xho</i> I (20 U/μL)	<i>Bam</i> HI e <i>Xho</i> I (10 U/μl)
	Excisão e Purificação	- Banda de peso ~1200 pb - <i>Kit Perfectprep Gel Cleanup®</i>	----- - <i>Colunas Illustra MicroSpin S300Hr®</i>
	Ligação	7 μL <i>insert</i> :1 μL vector; <i>Rapid DNA Ligation Kit</i>	
	Transformação	- Bactérias quimicamente competentes da estirpe TOP10 de <i>Escherichia coli</i> - Placa LB-Agar-Amp, 37°C	
	PCR selecção	<i>Primers</i> : T7 forward/M13 reverse	<i>Primers</i> : pCMV5 forward/pfGHR1 reverse
		Prog.: 95°C 10 min 94°C 45 s 54°C 45 s 72°C 1 min 30 s 72°C 15 min } x 35	Prog.: 95°C 10 min 94°C 45 s 58°C 45 s 72°C 1 min 30 s 72°C 15 min } x 35
	Mini-Cultura	- LB líquido + Ampicilina + Colónia(s) positiva(s) - Agitação constante (220 rpm) a 37°C, ~16h	
	Extracção de DNA Plasmídico	<i>peqGOLD Plasmid Miniprep Kit I</i>	
	Sequenciação	<i>Primers</i> : T7 forward/M13 reverse	<i>Primers</i> : T7 forward/ "IDS" reverse

3.2.6.5 Cultura de células para transfecção

Para poder proceder à transfecção do minigene construído, a linha celular de cancro colorectal DT-1 foi mantida em cultura. Esta linha foi estabelecida a partir de um adenocarcinoma colorectal em que as células apresentam uma morfologia epitelial aderente, desenvolvendo-se em monocamada.

A cultura da linha celular DT-1 iniciou-se com a descongelação rápida, a 37°C, de uma ampola de células (1 ml) conservada em azoto líquido, que após uma suave homogeneização foi semeada em 19 ml de DMEM [Gibco®] pré-aquecido, suplementado com 10% de FBS [Gibco®]. O frasco de cultura de células, de 80 cm² (T80) [Orange Scientific®], foi colocado em estufa de incubação a 37°C, numa atmosfera modificada com 5% de CO₂. Esta linha foi mantida nestas condições até se apresentar confluenta. Posteriormente, o meio de crescimento foi retirado e adicionada tripsina [Gibco®] de modo a dissociar as células da caixa de cultura. De seguida as células foram transferidas numa diluição adequada para um novo frasco T80 contendo meio fresco (DMEM+10%FBS). Para a transfecção dos plasmídeos foram usadas placas de 35mm de diâmetro (p35), nas quais foram semeadas cerca de 1x10⁶ células que foram depois incubadas durante a noite na estufa a 37°C, numa atmosfera modificada com 5% de CO₂. Após 24 horas verificou-se que as células se encontravam com cerca de 60-80% de confluência, permitindo a sua eficiente transfecção.

3.2.6.5.1 Transfecção

O minigene construído no vector pcDNA3-Myc1® para o doente foi transfectado em células DT-1 com cerca de 60-80% de confluência, utilizando o reagente de transfecção *Lipofectamine*TM 2000 (LF2K) [InvitrogenTM] segundo as instruções do fabricante. O DNA plasmídico a transfectar e o reagente *Lipofectamine*TM 2000 foram diluídos em meio Opti-MEM® I *Reduced Serum Medium* [InvitrogenTM] antes de serem colocados em contacto para formação dos complexos lipossoma-ácido nucleico.

As condições em que foi efectuada a transfecção e a quantidade de DNA plasmídico utilizado, encontram-se descritas na tabela VII.

Tabela VII. Condições e quantidades utilizadas na transfecção em células DT-1 do plasmídeo pcDNA3-Myc1[®] construído com a sequência mutada (c.257C>T) do gene *IDS*.

Condições de transfecção	Quantidade de plasmídeo
	pcDNA3-Myc1 [®] Doente
- 1x10 ⁶ células/placa 10cm ² (p35) - 125 µL OptiMEM [®] -DNA - 125 µL OptiMEM [®] -LF2K [®] - 4 µL LF2K [®]	2 µg

Abreviaturas: LF2K[®] - *Lipofectamine*TM 2000

3.2.6.5.2 Extracção de RNA total e síntese de cDNA após transfecção

Após a incubação da placa transfectada a 37°C, com 5% de CO₂, durante a noite (16-20h), procedeu-se à recolha das células para extracção de RNA total. O processo de extracção foi efectuado com o *kit High Pure RNA Isolation* [Roche] tal como descrito em 3.2.3.

Depois de obtido o RNA total da amostra, o respectivo cDNA foi sintetizado com o *kit First-Strand cDNA Synthesis* [Amersham Biosciences] conforme se descreve em 3.2.4.1.

3.2.6.5.3 Amplificação e sequenciação dos fragmentos de cDNA do minigene *IDS* obtidos após transfecção

Com o objectivo de analisar os fragmentos de cDNA do minigene *IDS* obtidos após a transfecção em células DT-1, foi efectuada a sua amplificação por PCR utilizando a mistura *Immomix Red 1x* [Bioline], os *primers* K-Myc-Bam *forward* e “IDS3” *reverse* com uma concentração de 0,5 µM e 5 µL de cDNA para um volume final de 25 µL (Tabela VIII).

Tabela VIII. Sequência dos *primers* utilizados na amplificação dos fragmentos de cDNA *IDS* obtidos após transfecção e respectivas condições de amplificação por PCR.

Sequência dos <i>primers</i> (5'→3')	Condições de PCR
F - GGTACCGCCACCATGGAGCAGAAGCTGATCTCCGAGGAGGACCTGGGATCC R - AAAGACTTTTCCACCGACA	95°C 7 min 94°C 45 seg Δ58-64°C 45 seg 72°C 45 seg 72°C 7 min <div style="display: flex; align-items: center; justify-content: center;"> <div style="font-size: 3em; margin-right: 10px;">}</div> <div style="text-align: center;"> 35 ciclos </div> </div>
Abreviaturas: F – <i>forward</i> ; R - <i>reverse</i>	

A amplificação dos fragmentos pretendidos foi verificada através da separação destes por electroforese em gel de agarose 2%, em tampão TAE 1x, juntamente com um marcador de pesos moleculares de 100pb [Fermentas]. Posteriormente efectuou-se a sequenciação dos fragmentos de interesse utilizando os *primers* descritos na tabela VIII e seguindo as condições descritas em 3.2.4.4.

Capítulo IV

Resultados e Discussão

4. Resultados e Discussão

4.1 Identificação e caracterização da mutação c.257C>T no gene *IDS* de um doente com Mucopolissacaridose II

Os métodos de análise mutacional utilizados e descritos no capítulo III, permitiram a identificação e caracterização da mutação c.257C>T (p.P86L) no gene *IDS* num doente com Mucopolissacaridose II (Síndrome de *Hunter*).

Para além da análise genómica, a análise por electroforese e sequenciação do cDNA do doente para o gene *IDS* com os *primers* 1F e 1R (que fazem *annealing* nos exões 1 e 5 respectivamente) representados na tabela I (Capítulo III) permitiu observar a ocorrência de dois transcritos na presença da referida mutação. Um dos transcritos apresentou-se como um produto normal de *splicing* codificando para uma proteína com a substituição aminoacídica P86L. Um segundo transcrito apresentou uma deleção dos primeiros 44 nucleótidos do exão 3 (Figura 11). Este transcrito dá origem a uma proteína bastante truncada com apenas 82 aminoácidos. Contudo, dado o aparecimento de um codão STOP prematuro, este transcrito será um potencial alvo do mecanismo de NMD.

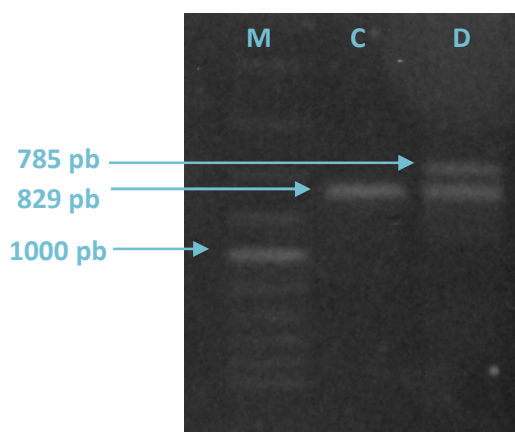


Figura 11: Análise do cDNA do doente com a mutação c.257C>T no gene *IDS*. C – cDNA do indivíduo controlo; D – cDNA do doente; M – marcador de pesos moleculares.

4.2 Análise Bioinformática

Os *scores* para os locais de *splicing* do intrão 2 foram calculados utilizando o *software* MaxEntScan (Yeo & Burge 2004) tendo-se obtido valores de *score* 9,80 para o 5' *splice-site* e 8,78 para o 3' *splice-site*. Valores estes relativamente elevados tendo em conta os *scores* de referência ideais deste programa – 11,81 (5'ss) e 13,59 (3'ss). Por outro lado, dada a ocorrência de um transcrito alternativo com a deleção dos primeiros 44 nucleótidos do exão 3, foi também efectuada a previsão do *score* para o local de *splicing* alternativo a 3' utilizado na formação deste transcrito. Para este local críptico de *splicing* o valor de *score* previsto foi de 8,89. Dado que este valor é mais elevado do que o valor encontrado para o local normal de *splicing* a 3', sugere que a utilização do local normal de *splicing* estará possivelmente dependente de elementos *cis-acting* e/ou factores *trans-acting* (Figura 12).

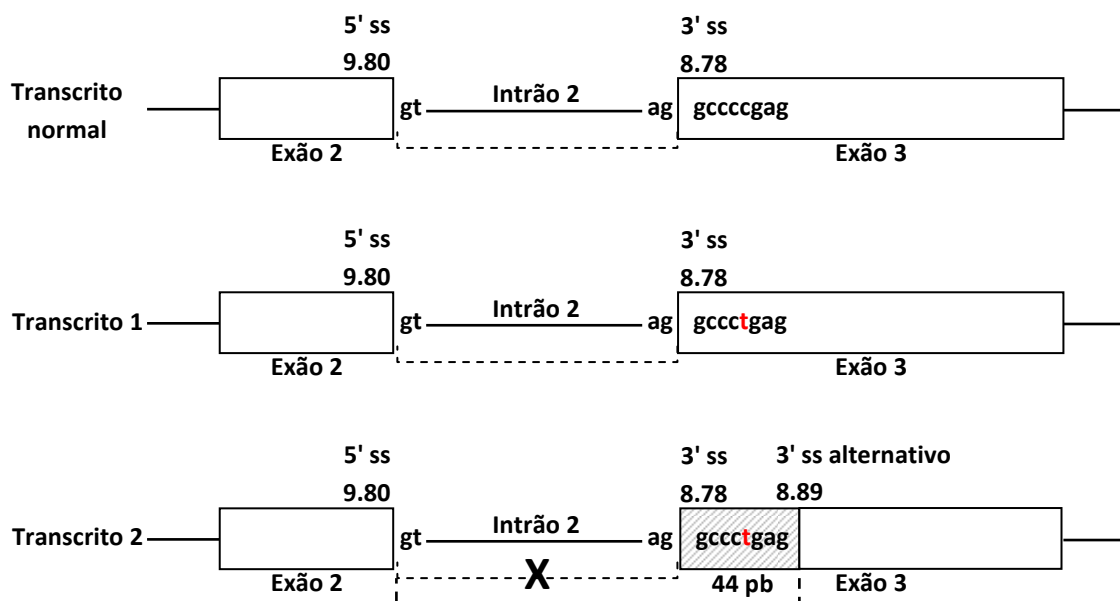


Figura 12: Representação esquemática dos transcritos obtidos na presença da mutação c.257C>T. Encontram-se também representados os diferentes valores de *score* para o 5' *splice-site* (ss), 3' *splice-site* (ss) e 3' *splice-site* (ss) alternativo.

Existem diversas mutações que induzem a utilização de locais de *splicing* aberrantes que podem ser classificados em duas categorias: (i) locais crípticos de *splicing*, usados quando a mutação interrompe o uso do local de *splicing*

constitutivo e (ii) locais de *splicing de novo*, que são criados por mutações ocorridas em qualquer local, quer dos intrões, quer dos exões provocando o aumento do reconhecimento e emparelhamento da maquinaria de *splicing* nesse local (Vorechovský 2006). Contudo, a distinção entre estas duas categorias pode ser um pouco ambígua em alguns casos, uma vez que a interrupção do local de *splicing* constitutivo poderá criar um novo local de *splicing*, sendo esta situação menos óbvia para os 3'ss (como é o nosso caso) do que para os 5'ss porque o reconhecimento preciso destes locais aceitadores requer a presença de sequências sinal localizadas no intrão. Estas sequências sinal, nomeadamente a BPS, o PPT e 3'AG, são reconhecidas por interacções RNA-proteína que envolvem a proteína SF1 e os snRNPs U2AF65 e 35. Assim, a “força” global dos 3'ss é definida pela presença de sequências óptimas necessárias à interacção com cada um dos factores mencionados, bem como pelas distâncias entre cada um deles aquando da sua ligação à respectiva sequência (Vorechovský 2006).

A activação de locais de *splicing* crípticos pode ser influenciada pelo balanço entre a “força” intrínseca do local de *splicing* aberrante e a “força” do local constitutivo, pela disponibilidade das sequências sinal localizadas na proximidade dos locais de *splicing* mutados, pelo tamanho dos exões e intrões, natureza da mutação e pela interrupção ou criação de ESEs, ESSs, ISEs ou ISSs. Contudo, embora se estime que o *skipping* exónico e a activação de locais de *splicing* crípticos/*de novo* sejam os principais acontecimentos que causam a vasta maioria dos produtos de *splicing* aberrantes transcritos a partir de alelos mutados, a extensão da contribuição dos elementos *cis-acting* para a activação de locais de *splicing* induzidos por uma mutação ainda não se encontra totalmente esclarecida (Královicová & Vorechovský, 2007).

A observação de que o *splicing* ocorre normalmente num local de *splicing* a 3' com um valor de *score* menor, sugere que a escolha deste 3'ss é efectuada com o auxílio de elementos *cis-acting* e/ou factores *trans-acting* que regulam o *splicing* naquele local e que de alguma forma foram alterados pela presença da mutação. Assim, uma vez obtidos estes dados, procedeu-se à previsão bioinformática de alterações em elementos *cis-acting* e *trans-acting* utilizando duas ferramentas

construídas com base em análises experimentais e estatísticas e que se encontram disponíveis na internet: *ESE finder 3.0* e *Splicing Rainbow*. O software *Splicing Rainbow* (Stamm *et al.*, 2006) permitiu identificar a eliminação de um local para a ligação das proteínas hnRNP E1 e E2 (Figura 13 A) e a criação de um motivo de ligação para a proteína SR - ASF/SF2 (Figura 13 B).

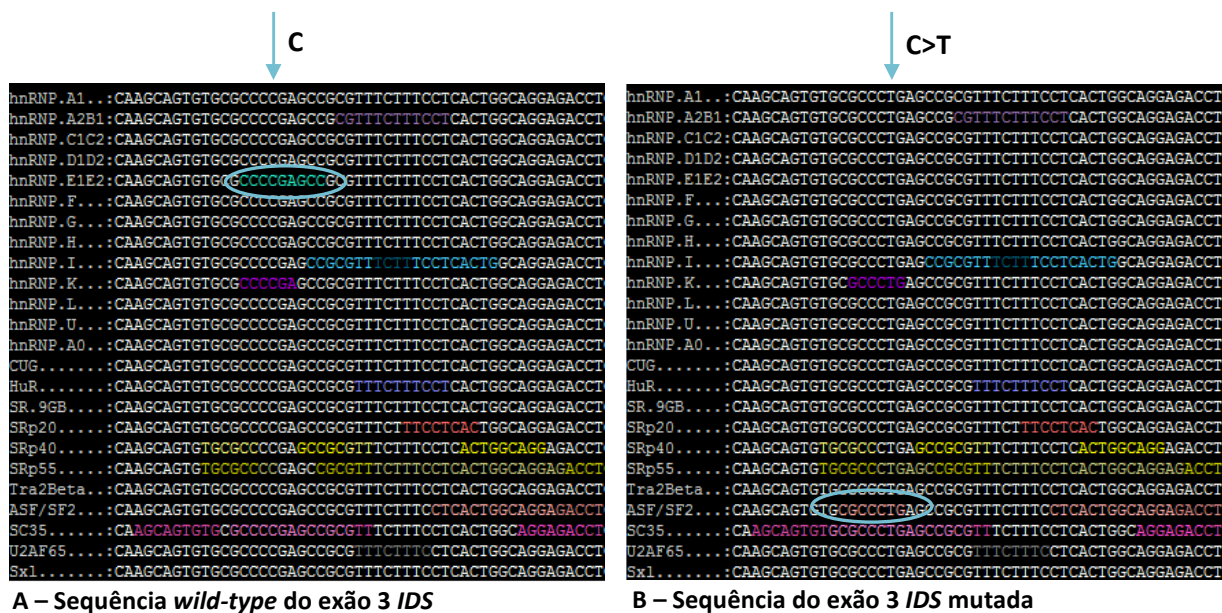


Figura 13: Representação das previsões efectuadas no software *Splicing Rainbow* para a sequência *wild-type* (A) e com a substituição nucleotídica C>T no exão 3 do gene *IDS* (B).

Por outro lado, a análise efectuada no programa *ESE finder 3.0* (Cartegni *et al.*, 2003) prevê que a proteína ASF/SF2 se ligue no caso da sequência *wild-type* a um motivo de um elemento *Exonic Splicing Enhancer* com a sequência, GCCCCGA que apresenta um valor de *score* de 2,29278 (*threshold* – 1,956) (Figura 14). Na presença da mutação, este local de ligação não é eliminado, mas sofre modificações. A sequência identificada como local de ligação da proteína ASF/SF2 é ligeiramente diferente, CGCCCTG e apresenta um *score* de 2,30151 (Figura 14). Curiosamente, este motivo previsto para a ligação da proteína ASF/SF2 no caso da sequência exónica com a mutação, é o mesmo motivo de ligação previsto pelo software *Splicing Rainbow*. Contudo, este não prevê qualquer ligação desta proteína no caso da sequência *wild-type*.

O programa *ESE finder 3.0* permitiu ainda prever que a substituição nucleotídica C>T elimina um putativo local de ligação para a proteína SR - SC35 com a sequência GCCCCGAG (Figura 15). No caso do *software Splicing Rainbow*, tanto a sequência *wild-type* como a sequência com a mutação possuem a previsão de um local de ligação para esta proteína.

Proteína ASF/SF2

SF2/ASF threshold: 1.956		
Position*/Site/Score		
13 (-166)	GCCCCGA	2.29278
34 (-145)	CTCACTG	2.61975
40 (-139)	GGCAGGA	3.96044
52 (-127)	GACACCA	2.43404
57 (-122)	CACCCGC	3.38890
61 (-118)	CGCCTGT	2.27135
96 (-83)	CGCTGGA	3.08688
119 (-60)	CCCAGTA	2.13949
155 (-24)	CGGTGGG	2.11387
172 (-7)	CACCCCTG	2.74873

Sequência *wild-type*

SF2/ASF threshold: 1.956		
Position*/Site/Score		
12 (-167)	CGCCCTG	2.30151
34 (-145)	CTCACTG	2.61975
40 (-139)	GGCAGGA	3.96044
52 (-127)	GACACCA	2.43404
57 (-122)	CACCCGC	3.38890
61 (-118)	CGCCTGT	2.27135
96 (-83)	CGCTGGA	3.08688
119 (-60)	CCCAGTA	2.13949
155 (-24)	CGGTGGG	2.11387
172 (-7)	CACCCCTG	2.74873

Sequência com a mutação
c.257C>T

Figura 14: Representação das previsões efectuadas no *software ESE finder 3.0*. Previsão efectuada para a proteína SR – ASF/SF2; posição exónica, motivo e *score* para a sequência *wild-type* e com a mutação c.257C>T.

Proteína SC35

SC35 threshold: 2.383		
Position*/Site/Score		
13 (-166)	GCCCCGAG	2.47716
70 (-109)	GACTTCAA	3.62772
76 (-103)	AACTCCTA	3.65800
93 (-86)	GCACGCTG	2.50722
115 (-64)	ATCCCCCA	2.95757

Sequência *wild-type*

SC35 threshold: 2.383		
Position*/Site/Score		
70 (-109)	GACTTCAA	3.62772
76 (-103)	AACTCCTA	3.65800
93 (-86)	GCACGCTG	2.50722
115 (-64)	ATCCCCCA	2.95757

Sequência com a mutação
c.257C>T

Figura 15: Representação das previsões efectuadas no *software ESE finder 3.0*. Previsão efectuada para a proteína SR – SC35; posição exónica, motivo e *score* para a sequência *wild-type* e com a mutação c.257C>T.

Os resultados obtidos pela utilização dos dois *softwares* levaram-nos a colocar algumas hipóteses explicativas para a regulação do *splicing* nesta região do gene *IDS* e de como este é alterado na presença da mutação c.257C>T.

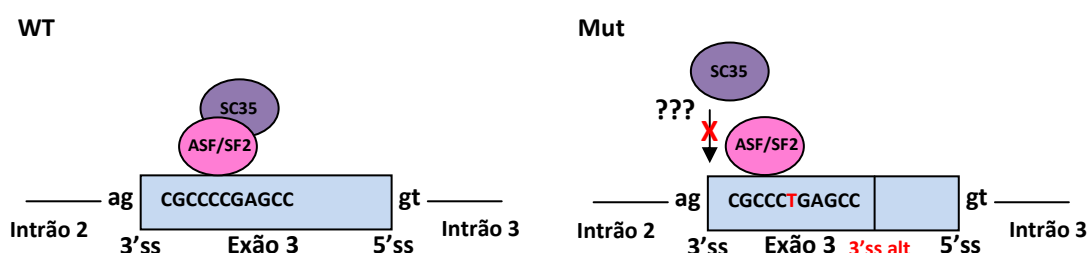
A análise pelo *ESEfinder*, que prevê a eliminação de um ESE para a proteína SC35, indica que esta proteína SR poderá estar envolvida na utilização do local de *splicing* normal uma vez que este apresenta um valor de *score* inferior ao local crítico de *splicing* (Figura 16 A).

Já os resultados obtidos pelo *software Splicing Rainbow*, que prevê a eliminação de um local de ligação para as proteínas hnRNP E1 e E2 na sequência mutada, com a consequente utilização de um local crítico de *splicing* no interior do exão 3, sugerem que estas proteínas possam estar envolvidas na repressão deste local crítico de *splicing* a 3'. Em simultâneo este programa prevê também a criação de um local de ligação para a proteína ASF/SF2 na sequência com a mutação, sugerindo que a ligação desta proteína SR poderá potenciar a utilização do local crítico de *splicing* no exão 3 sobretudo se ao mesmo tempo ocorrer a perda do local de ligação para as proteínas hnRNP E1 e E2 (Figura 16 B).

De facto, a regulação do local de *splicing* a 3' do intrão 2 parece ser bastante complexa, e a presença desta mutação terá vindo perturbar em parte (já que o

transcrito normal ainda continua a ser produzido) o equilíbrio entre os diversos elementos *cis* e factores *trans-acting* que nela participam. A utilização do local normal de *splicing* em detrimento do local críptico (apesar de este apresentar um valor de *score* superior) parece depender da interacção de várias proteínas (quer activadoras quer silenciadoras), da sua ligação a determinadas sequências de DNA e também da sua interacção com os elementos do spliceossoma.

A) *ESE finder*



B) *Splicing Rainbow*

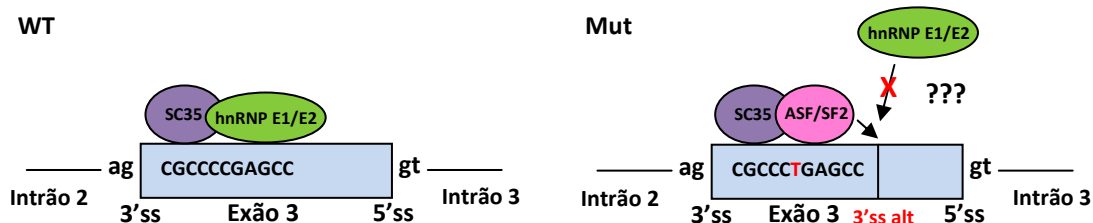


Figura 16: Modelo proposto para a regulação do *splicing* do exão 3 do gene *IDS* com base nas previsões efectuadas pelo *software* *ESE finder* (A) e *Splicing Rainbow* (B).

As proteínas ASF/SF2 e SC35 pertencem à família “clássica” das proteínas SR que compreende um grupo de proteínas conservadas filogeneticamente e relacionadas estruturalmente que possuem uma ou duas cópias de um *RNA-recognition motif* (RRM) e um domínio característico rico em resíduos arginina-serina, conhecido como domínio RS. As proteínas pertencentes a esta família desempenham um papel importante tanto no processo de *splicing* constitutivo do pré-mRNA, como na regulação do *splicing* alternativo, através do reconhecimento de elementos *cis-acting* (activadores), influenciando a escolha dos locais de *splicing*

através da interacção com snRNPs (U1; U2AF65/35), como é o caso das proteínas ASF/SF2 e SC35. Estas proteínas exercem ainda funções em algumas actividades que ocorrem após o processo de *splicing*, como a exportação nuclear do mRNA, o mecanismo de NMD e a própria tradução do mRNA. Esta diversidade de funções das proteínas SR realça a sua importância como reguladores essenciais do metabolismo do mRNA (Manley & Tacke, 1996; Caceres & Kornblitt 2002; Long & Caceres, 2009). A análise de cerca de 50 substituições nucleotídicas associadas ao *skipping* de exões em genes humanos revelou que mais de 50% destas mutações anulava pelo menos um motivo alvo para as proteínas SR – ASF/SF2, SRp40, SRp55 e SC35 (Cartegni *et al.*, 2002). Deste modo, qualquer alteração ocorrida que modifique uma das muitas funções das proteínas SR poderá estar associada ao aparecimento de doença.

Por outro lado, as proteínas hnRNP, entre as quais a hnRNP E1 e E2 previstas pelo *software Splicing Rainbow* como elementos repressores envolvidos no *splicing* do exão 3 do gene *IDS*, são proteínas que possuem na sua estrutura um domínio de ligação ao RNA, associado a um domínio auxiliar normalmente envolvido em interacções proteína-proteína. De um modo geral as proteínas hnRNP possuem funções em várias actividades celulares, desde a transcrição e o processamento do pré-mRNA no núcleo até ao processo de tradução no citoplasma (Krecic & Swanson, 1999).

Especificamente as proteínas hnRNP E1 e E2 pertencem ao grupo das proteínas hnRNP E constituído por 5 isoformas – hnRNP E1 a E4 e hnRNP E2-KL, uma variante de *splicing* da proteína hnRNP E2 que difere da original por uma perda de 31 aminoácidos na região entre os domínios KH2 e KH3 (Makeyev & Liebhaber, 2000). As duas isoformas E1 e E2 são as formas proteicas habitualmente com maior expressão e as mais bem caracterizadas, possuindo 89% de homologia a nível aminoacídico (Chkheidze & Liebhaber, 2003). A nível funcional encontram-se envolvidas na estabilização do mRNA e no silenciamento da tradução (Krecic & Swanson, 1999; Woolaway *et al.*, 2007). Tal como no caso das proteínas SR, qualquer alteração que implique a perda de função destas proteínas poderá ter repercussões negativas ao nível do normal funcionamento do processo de *splicing* podendo também estar na origem de algumas doenças.

Algumas proteínas da família SR que se ligam habitualmente a ESEs, como é o caso da ASF/SF2, podem actuar também como antagonistas da actividade repressora das proteínas hnRNP que se ligam a ESSs, influenciando assim a escolha dos locais de *splicing*. Deste modo, o *ratio* entre as proteínas SR e hnRNP existente no núcleo assume elevada importância na regulação do *splicing* alternativo (Caceres & Kornblitt 2002; Long & Caceres, 2009).

A importância destas proteínas na regulação do processo de *splicing* e as previsões obtidas de certo modo discordantes entre os dois *softwares* bioinformáticos, reforçam a importância de ser efectuada uma análise funcional do *splicing* no gene *IDS* na presença e na ausência da mutação c.257C>T através do uso de minigenes, assim como a sobreexpressão e silenciamento (RNAi) destas proteínas para poder avaliar a exactidão das previsões bioinformáticas e conseguir validar alguma das nossas hipóteses.

4.3 Análise funcional do impacto da mutação c.257C>T

4.3.1 Obtenção de minigenes repórter intermédios

Para estudar funcionalmente a mutação c.257C>T foram construídos num primeiro passo minigenes repórter intermédios no vector plasmídico pCR2.1®-TOPO® [Invitrogen™] para um controlo e doente.

Os fragmentos de gDNA do controlo e doente com 1137 pb foram amplificados por PCR em gradiente de temperaturas (Figura 17). No caso do gDNA do controlo foi utilizada para purificação a amostra amplificada a 61°C e no caso do doente a amostra amplificada a 59°C.

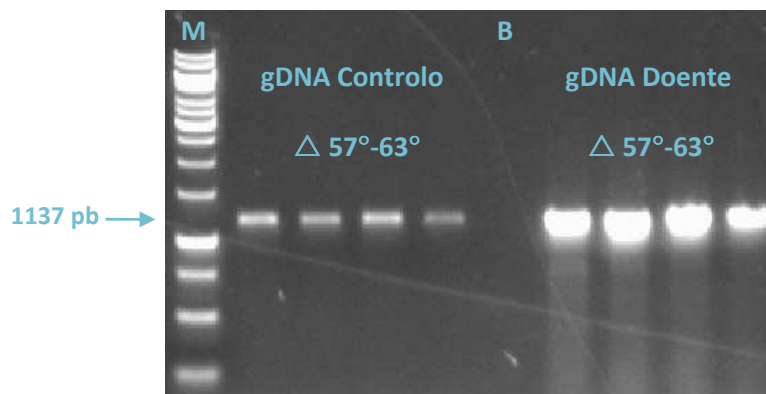


Figura 17: Amplificação em gradiente dos fragmentos de interesse do gene *IDS* para o controlo e doente. gDNA (DNA genómico) Controlo; gDNA (DNA genómico) Doente; $\Delta 57^{\circ}$ - 63° (Intervalo de temperaturas de amplificação - 57°C , 59°C , 61°C e 63°C); B – branco; M – marcador de pesos moleculares.

Após clonagem no vector pCR2.1®-TOPO® [Invitrogen™] e reacção de transformação em bactérias quimicamente competentes, a reacção de PCR de selecção permitiu identificar tanto para o controlo como para o doente, colónias positivas, isto é colónias de bactérias que integraram plasmídeos com os fragmentos de interesse (1137pb), assim como uma colónia azul (controlo negativo) indicadora da ausência de inserção do fragmento de interesse (Figura 18 A e B).

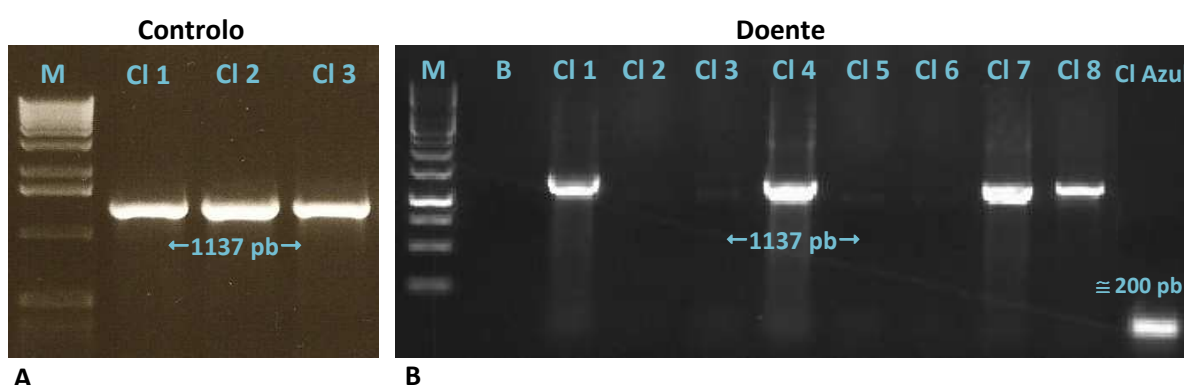


Figura 18: Selecção por PCR de colónias positivas (plasmídeos pCR2.1®-TOPO® [Invitrogen™] com fragmentos de interesse inseridos) para o controlo (A) e doente (B). CI – Colónia; B – branco; M – marcador de pesos moleculares.

Depois da selecção por PCR, foi efectuada a sequenciação das colónias positivas escolhidas tanto para o caso controlo (Cl2) como para o doente (Cl8). No que se refere à colónia 2 do caso controlo, a sequenciação permitiu verificar que o fragmento de interesse clonado no vector pCR2.1®-TOPO® [Invitrogen™] continha duas substituições nucleotídicas inseridas, uma no exão 2 (A>G) e outra no intrão 2 (T>C) (Figura 19 A e B). Estas substituições, uma vez presentes na sequência do fragmento de interesse poderão alterar elementos *cis-acting* necessários ao normal funcionamento da maquinaria de *splicing*, invalidando assim o seu uso posterior para construção do minigene final no vector pcDNA3-Myc1®. O aparecimento destas substituições de nucleótidos no fragmento terá ocorrido possivelmente durante o processo de amplificação na reacção de PCR onde foi utilizada a mistura *ImmoMix Red 1x* [Bioline] que contém a enzima DNA polimerase *Immolase*. Esta mistura apesar de permitir efectuar os ensaios em menor tempo, com menor risco de contaminação, com elevada especificidade e reprodutibilidade e de possuir aplicação para o processo de clonagem (adição de 3'-A *overhangs*), não possibilitou a obtenção de fragmentos de interesse sem qualquer alteração na sequência. Para contornar esta situação deverá ser utilizada uma enzima com “proofreading” como por exemplo a DNA polimerase *Pfu* ou *Taq Plus Precise*. Qualquer uma destas enzimas permitirá realizar a amplificação de DNA a clonar com elevada exactidão de modo a que não sejam introduzidas substituições nucleotídicas nos fragmentos de interesse (Cooper 2005).

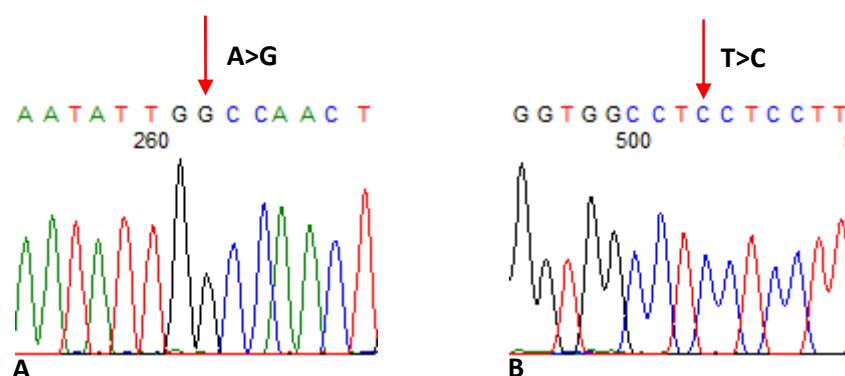


Figura 19: Identificação das substituições nucleotídicas ocorridas na sequência de gDNA do fragmento controlo por sequenciação automática. (A) Sequência parcial do gDNA clonado que permite identificar a substituição A>G no exão 2 do gene *IDS*. (B) Sequência parcial do gDNA clonado onde se pode observar a substituição T>C no intrão 2 do gene *IDS*.

Relativamente à colónia 8 escolhida no caso do doente, foi possível observar através da sequenciação que o fragmento clonado se encontrava íntegro e com a alteração nucleotídica C>T no exão 3, não tendo ocorrido portanto nenhuma alteração.

4.3.2 Obtenção do minigene repórter final construído com a sequência mutada (c.257C>T) do gene *IDS*

Dado que o minigene repórter intermédio para o caso controlo possuía alterações que invalidavam o seu uso para a construção do minigene repórter final, apenas se procedeu à construção do minigene repórter final para o fragmento de interesse do doente em estudo.

No primeiro passo desta construção foi efectuada a digestão enzimática (*Bam*HI [NEB®_{inc}] e *Xho*I [NEB®_{inc}]) do vector pcDNA3-Myc1® e do minigene intermédio contendo clonado o fragmento de gDNA *IDS* do doente. Na figura 20 A e B pode ser verificado o sucesso destas digestões. Na linha 1 da figura 20 A é possível visualizar o plasmídeo pcDNA3-Myc1® digerido, uma vez que se apresenta com o peso molecular (5.4 Kb) referente à sua forma linearizada. Na linha 2 encontra-se representado o plasmídeo não digerido utilizado apenas como amostra controlo para comparação com o plasmídeo digerido. Na figura 20 B pode ser observada na linha 1 a digestão do minigene intermédio. A seta vermelha indica o fragmento de gDNA de interesse com 1137 pb dissociado do vector pcDNA3-Myc1® representado pela seta verde (5.4 Kb).

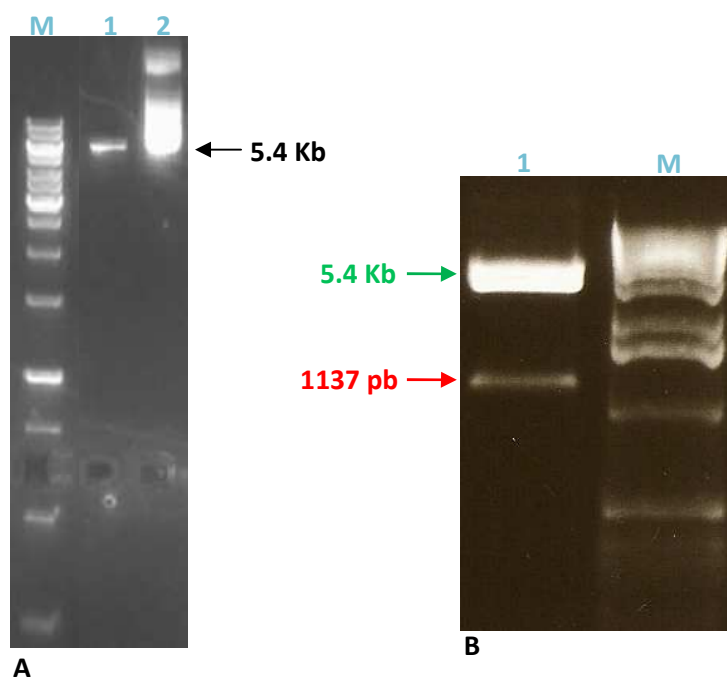


Figura 20: Digestão enzimática do vector pcDNA3-Myc1® e do minigene repórter intermédio contendo clonado o fragmento de gDNA *IDS* do doente com as enzimas de restrição *Bam*HI® e *Xho*I®. (A) Linha 1 - vector pcDNA3-Myc1® digerido; Linha 2 - vector pcDNA3-Myc1® não digerido. (B) Linha 1 – digestão do minigene intermédio; Seta vermelha – fragmento de interesse digerido; Seta verde – vector pcDNA3-Myc1® digerido. M – marcador de pesos moleculares.

Depois de digeridas as amostras, ligadas e transformadas em bactérias quimicamente competentes foi efectuada a reacção de PCR de selecção que possibilitou a identificação de duas colónias positivas, isto é, colónias de bactérias que integraram o vector pcDNA3-Myc1® com a fracção de gDNA *IDS* do doente inserida (Figura 21).

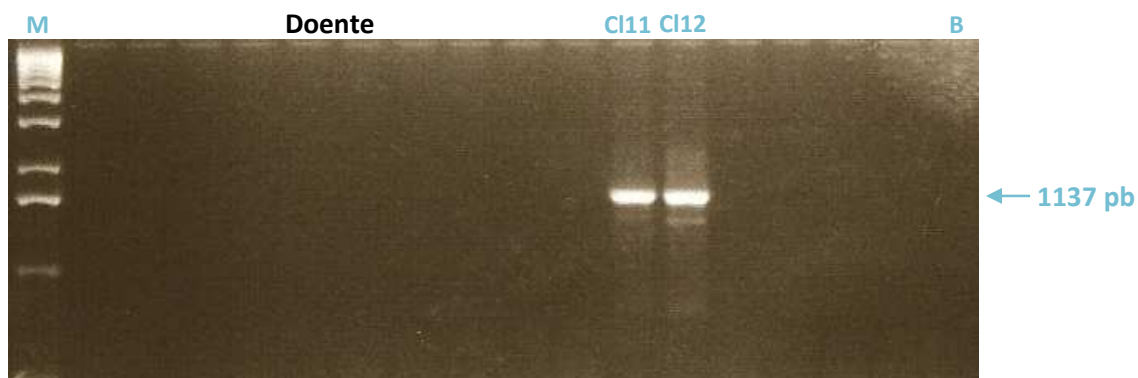


Figura 21: Selecção por PCR de colónias positivas (pcDNA3-Myc1® [Invitrogen™] com o fragmento de interesse do doente inserido). Colónias positivas – número 11 e 12; B – branco; M – marcador de pesos moleculares.

Após a selecção, a sequenciação das duas colónias positivas permitiu verificar que tanto num caso como noutro o fragmento de interesse se encontrava inserido correctamente no plasmídeo pcDNA3-Myc1®, sendo possível verificar na sequência a alteração nucleotídica C>T no exão 3. Como exemplo é apresentada na figura 22 B, a sequenciação no local da mutação para a colónia 11. Esta colónia foi escolhida para posteriormente ser extraído, por *kit* miniprep, o DNA plasmídico correspondente ao minigene final. A figura 22 A representa a sequência normal do mesmo fragmento do gDNA *IDS*, tendo sido obtida no entanto a partir da sequenciação de gDNA de um indivíduo controlo. Esta figura é exibida apenas para que seja possível comparar a sequência *wild-type* com a mutante do doente em estudo, uma vez que não se prosseguiu com a obtenção do minigene final para o caso controlo.

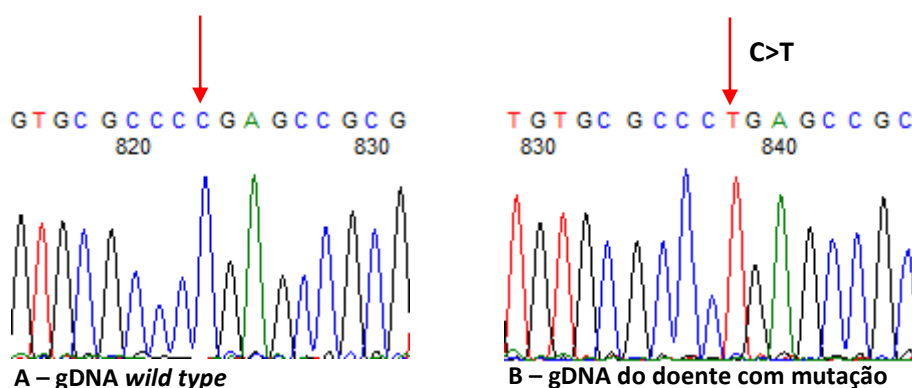


Figura 22: Representação das sequências do gDNA *wild-type* e gDNA do doente em estudo com a mutação c.257C>T no gene *IDS*. (A) Sequência normal do fragmento de interesse do gene *IDS* obtida de um controlo. (B) Sequência do fragmento de interesse mutado pertencente ao doente, obtida após inserção no vector pcDNA3-Myc1®.

Depois de obtida a miniprep com o plasmídeo contendo o gDNA de interesse do doente, a amostra foi quantificada espectrofotometricamente e apresentou uma concentração de 325 ng/μL e um *ratio* (A_{260}/A_{280}) de 1,98, encontrando-se este dentro do valor aceitável para o parâmetro contaminação por proteínas. Dado que a concentração se apresentou maior do que o valor mínimo normalmente necessário (200 ng/μL) e o grau de contaminação proteica dentro do valor aceite, foi possível utilizar esta amostra para transfecção nas células DT-1.

4.3.3 Transfecção transiente do minigene repórter final (pcDNA3-Myc1® + fragmento gDNA *IDS* doente)

Uma vez obtido o minigene final (Figura 23) construído com o fragmento de gDNA de interesse do doente procedeu-se à sua transfecção em células DT-1. O objectivo desta transfecção era avaliar se o minigene contendo o fragmento de gDNA *IDS* com a mutação de *splicing* c.257C>T, reproduzia *in vivo*, os mesmos resultados observados aquando da análise do cDNA do doente.

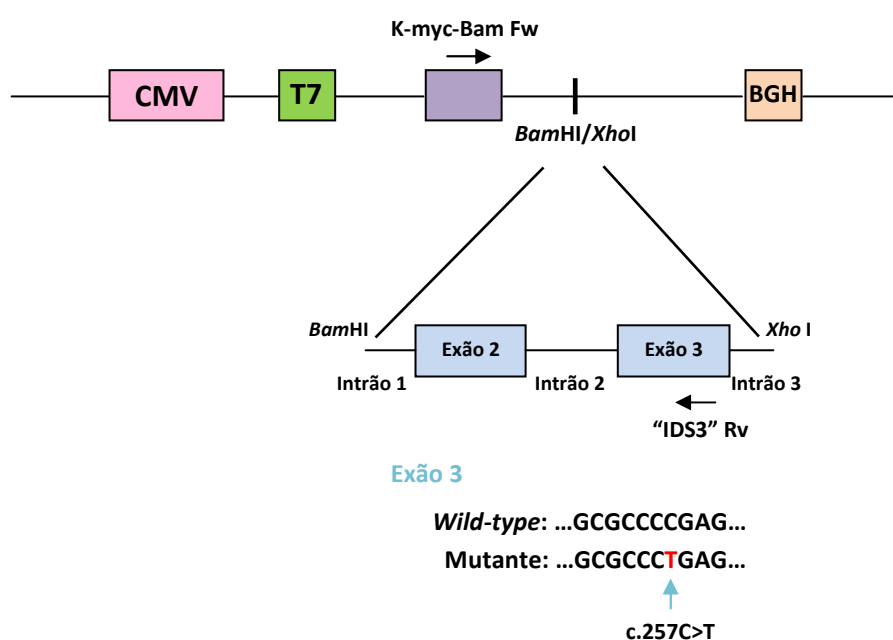


Figura 23: Representação esquemática do minigene repórter final construído com o fragmento de interesse do doente no vector plasmídico pcDNA3®-Myc. Encontram-se representados os *primers* (K-myc-Bam *forward* e “IDS3” *reverse*) utilizados para amplificação do fragmento após ocorrência do *splicing in vivo*.

Após o processo de transfecção, foi então extraído o RNA das células, sintetizado o cDNA e realizada a reacção de PCR para amplificação dos fragmentos de cDNA do minigene *IDS* final. Nesta amplificação eram esperados, tal como observado no cDNA *in vitro*, dois transcritos, um maioritário correspondente ao exão 2 e 3 e outro de menor tamanho relativo ao exão 2 e exão 3 sem os primeiros 44 nucleótidos da sua sequência (Figura 24).

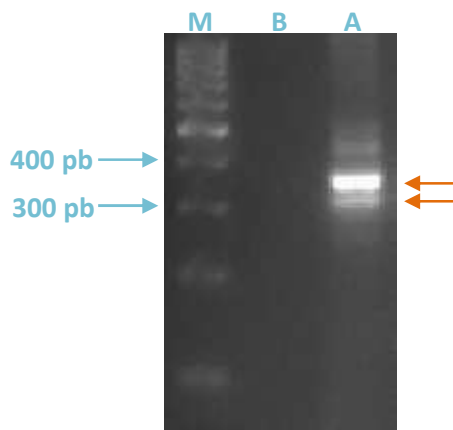


Figura 24: Amplificação do cDNA do fragmento de interesse do gene *IDS* do doente, obtido após ocorrência do processo de *splicing* no minigene repórter final transfectado em células DT-1. As setas cor de laranja indicam a presença dos dois transcritos. M – marcador de pesos moleculares; B – branco; A – amostra.

A posterior sequenciação destes fragmentos permitiu confirmar que o *splicing* no minigene ocorreu de acordo com o esperado (isto é da mesma forma que no doente), sugerindo que o minigene obtido é um sistema válido que poderá ser utilizado para a identificação de elementos *cis* e factores *trans* envolvidos na regulação do *splicing* desta região do gene *IDS*.

Capítulo V

Conclusão e Perspectivas Futuras

5. Conclusão e Perspectivas Futuras

A importância do processo de *splicing* ocorrer correctamente é evidenciada pelo facto de entre 15-60% das mutações associadas a doença estarem relacionadas com alterações, quer nas sequências consenso dos locais de *splicing*, quer nos elementos exónicos e intrónicos auxiliares (activadores ou repressores) envolvidos na regulação do processo de *splicing*. Deste modo, as mutações relacionadas com doenças constituem uma base importante para a realização de estudos funcionais que permitam clarificar tanto o seu efeito no *splicing* como contribuir para o conhecimento dos mecanismos moleculares básicos envolvidos neste processo de um modo geral.

O presente estudo tinha como principal objectivo a implementação laboratorial da técnica de clonagem e construção de minigenes para analisar funcionalmente o efeito da mutação c.257C>T no processo de *splicing* do gene *IDS*.

Através da análise do cDNA do doente foi possível observar na presença desta mutação a ocorrência de dois transcritos alternativos. Um dos transcritos apresentou-se como um produto normal de *splicing* codificando para uma proteína com a substituição do aminoácido prolina por leucina ao nível do codão 86. Um segundo transcrito apresentou uma deleção que envolvia os primeiros 44 nucleótidos do exão 3 dando origem a uma proteína bastante truncada.

A análise bioinformática para obter os *scores* dos locais de *splicing* permitiu prever que o local de *splicing* críptico a 3' utilizado na presença da mutação e que originou o transcrito alternativo com menos 44 nucleótidos possui um valor de *score* mais elevado (8,89) do que o local de *splicing* constitutivo (8,78). A partir desta observação foi possível colocar a hipótese de que a escolha deste local de *splicing* constitutivo (com um valor de *score* menor) poderá dever-se a um mecanismo de regulação nesse local por parte de elementos *cis-acting* e factores *trans-acting*. Dada esta hipótese e após o estudo bioinformático concluiu-se que alguns motivos *cis-acting* foram alterados na presença da mutação c.257C>T com consequente alteração das proteínas tanto activadoras (ASF/SF2 e SC35) como

repressoras (hnRNP E1 e E2) que se ligam especificamente a cada um desses motivos.

Foi então, efectuada a clonagem e construção de um minigene com 1137 pb contendo parte do intrão 1, o exão 2, intrão 2, exão 3 e parte do intrão 3, tanto a partir da amostra do doente com esta mutação como de um controlo. Foi possível, no caso do doente, obter o constructo final e após transfecção em células DT-1 verificar que o processo de *splicing* ocorreu de acordo com o esperado, sugerindo que este minigene será apropriado para o estudo de elementos *cis-acting* e factores *trans-acting* envolvidos na regulação do *splicing* nesta região do gene *IDS*.

Este trabalho permitiu ainda a implementação da técnica de clonagem e construção de minigenes no grupo de investigação em doenças lisossomais de sobrecarga do CGM-INSA, abrindo assim a possibilidade de continuar a efectuar estudos funcionais que contribuam para o conhecimento dos mecanismos de *splicing* em geral e de delinear novas abordagens terapêuticas nesta área.

Dado que durante a elaboração deste trabalho não foi possível efectuar a transfecção do constructo com a sequência *wild-type*, uma vez que apresentava substituições nucleotídicas que poderiam interferir com o processo de *splicing* da região de interesse, a sua obtenção será um objectivo a realizar a curto prazo para que possa servir como referência no estudo funcional do processo de *splicing* na referida região do gene *IDS*.

Posteriormente, outro objectivo será efectuar a sobreexpressão e/ou o silenciamento das proteínas ASF/SF2, SC35 e hnRNP E1 e E2 cujos motivos de ligação, segundo as previsões bioinformáticas foram alterados pela presença da mutação c.257C>T, através da sua co-expressão individual com cada um dos minigenes construídos para o controlo e doente para então avaliar o seu papel no *splicing* do exão 3 do gene *IDS*.

Capítulo VI

Bibliografía

6. Bibliografia

Almeida MR, Saraiva MJ. Parte 2 – Aplicações. Capítulo IX – Doenças Genéticas Humanas. Em Videira A. Editor. Engenharia Genética – Princípios e Aplicações. Lidel; 2001. p. 93 a 108.

Alter J, Lou F, Rabinowitz A, Yin H, Rosenfeld J, Wilton SD, Partridge TA, Lu QL. *Systemic delivery of morpholino oligonucleotide restores dystrophin expression bodywide and improves dystrophic pathology*. Nature medicine 2006; 12(2):175-177.

Alves S, Mangas M, Prata MJ, Ribeiro G, Lopes L, Ribeiro H, Pinto-Basto J, Lima MR, Lacerda L. *Molecular characterization of Portuguese patients with mucopolysaccharidosis type II shows evidence that the IDS gene is prone to splicing mutations*. Journal of Inherited Metabolic Disorders 2006; 29:743-754.

Ballabio A, Gieselmann V. *Lysosomal disorders: from storage to cellular damage*. Biochimica et Biophysica Acta 2009; 1793(4):684-696.

Ben-Dov C, Hartmann B, Lundgren J, Valcárcel J. *Genome-wide Analysis of Alternative Pre-mRNA Splicing*. Journal of Biological Chemistry 2008; 283(3):1229-1233.

Berget SM. *Exon recognition in vertebrate splicing*. The journal of biological chemistry 1995; 270(6):2411-2414.

Berget SM, Moore C, Sharp PA. *Spliced segments at the 5'terminus of adenovirus 2 late mRNA*. Proceedings of the National Academy of Sciences of the United States of America 1977; 74:3171-3175.

Black DL. *Mechanisms of alternative pre-messenger RNA splicing*. Annual Review of Biochemistry 2003; 72:291-336.

Blencowe BJ. *Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases*. Trends in Biochemical Sciences 2000; 25(3):106-110.

Blencowe BJ. *Alternative splicing: new insights from global analyses*. Cell 2006; 126(1):37-47.

Buratti E, Baralle FE. *Influence of RNA secondary structure on the pre-mRNA splicing process*. Molecular and cellular biology 2004; 24(24):10505-10514.

Burge C, Karlin S. *Prediction of complete gene structures in human genomic DNA*. Journal of molecular biology 1997; 268:78-94.

Cáceres JF, Kornblihtt AR. *Alternative splicing: multiple control mechanisms and involvement in human disease*. Trends in Genetics 2002; 18(4):186-193.

Cartegni L, Krainer AR. *Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1*. Nature Genetics 2002; 30:377-384.

Cartegni L, Chew SL, Krainer AR. *Listening to silence and understanding nonsense: exonic mutations that affect splicing*. Nature Reviews Genetics 2002; 3:285-298.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. *ESEfinder: A web resource to identify exonic splicing enhancers*. Nucleic Acids Research 2003; 31(13):3568-3571.

Cartegni L, Hastings ML, Calarco JA, de Stanchina E, Krainer AR. *Determinants of Exon 7 Splicing in the Spinal Muscular Atrophy Genes SMN1 and SMN2*. The American Journal of Human Genetics 2006; 78:63-77.

Chiara MD, Reed R. *A two-step mechanism for 5' and 3' splice-site pairing*. Nature 1995; 375(6531):510-513.

Chkheidze AN, Liebhaber SA. *A novel set of nuclear localization signals determine distributions of the alphaCP RNA-binding proteins*. Molecular & Cellular Biology 2003; 23:8405-8415.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. *An amazing sequence arrangement at the 5'ends of adenovirus 2 messenger RNA*. Cell 1977; 12:1-8.

Cooper GM, Hausman RE. Part II - The flow of genetic information. Chapter 5 - *The Organization and Sequences of Cellular Genomes*. In *The Cell: A Molecular Approach* 5th edition. Sinauer Associates Inc., Sunderland, Massachusetts: ASM Press Washington D.C.; 2009. p. 326 a 342.

Cooper TA. *Use of minigene systems to dissect alternative splicing elements*. Methods 2005; 37:331-340.

Cooper TA, Wan L, Dreyfuss G. *RNA and Disease*. Cell 2009; 136:777-793

Corcos L, Solier S. *Alternative mRNA splicing, pathology and molecular therapeutics*. Medecine/Sciences 2005; 21:253-260.

Cramer P, Pesce CG, Baralle FE, Kornblihtt AR. *Functional association between promoter structure and transcript alternative splicing*. Proceedings of the National Academy of Sciences of the United States of America 1997; 94(21):11456-60.

de la Mata M, Kornblihtt AR. *RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20*. Natural Structure & Molecular Biology 2006; 13(11):973-980.

de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR. *A slow RNA polymerase II affects alternative splicing in vivo*. Molecular Cell 2003; 12(2):525-532.

Disset A, Bourgeois CF, Benmalek N, Claustres M, Stevenin J, Tuffery-Giraud S. *An exon skipping-associated nonsense mutation in the dystrophin gene uncovers a complex interplay between multiple antagonistic splicing elements*. Human Molecular Genetics 2006; 15(6): 999-1013.

Donahue CP, Muratore C, Wu JY, Kosik KS, Wolfe MS. *Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing*. The journal of biological chemistry 2006; 281(33):23302-23306.

Dredge BK, Darnell RB. *Nova regulates GABA(A) receptor gamma2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer*. Molecular and cellular biology 2003; 23(13):4687-4700.

Faustino NA, Cooper TA. *Pre-mRNA splicing and human disease*. Genes & Development 2003; 17:419-437.

Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. *The architecture of pre-mRNAs affects mechanisms of splice-site pairing*. Proceedings of the National Academy of Sciences of the United States of America 2005; 102(45): 16176-16181.

Futerman AH, van Meer G. *The Cell Biology of Lysosomal Storage Disorders*. Nature Reviews Molecular Cell Biology 2004; 5:554-565.

Garcia-Blanco MA. *Methods for the study of alternative splicing*. Methods 2005; 37:289-291.

Garcia-Blanco MA, Baraniak AP, Lasda EL. *Alternative splicing in disease and therapy*. Nature Biotechnology 2004; 22(5):535-546.

Gonçalves V, Theisen P, Antunes O, Medeira A, Ramos JS, Jordan P, Isidro G. *A missense mutation in the APC tumor suppressor gene disrupts an ASF/SF2 splicing enhancer motif and causes pathogenic skipping of exon 14*. Mutation research 2009; 662:33-36.

Hastings ML, Krainer AR. *Pre-mRNA splicing in the new millennium*. Current Opinion in Cell Biology 2001; 13:302-309.

Hertel KJ. 2008. *Combinatorial Control of Exon Recognition*. Journal of Biological Chemistry 2008; 283(3):1211-1215.

Horowitz DS, Krainer AR. *Mechanisms for selecting 5'splice sites in mammalian pre-mRNA splicing*. Trends in Genetics 1994; 10(3):100-106.

House AE, Lynch KW. *Regulation of Alternative Splicing: More than Just the ABCs*. Journal of Biological Chemistry 2008; 283(3):1217-1221.

Hua Y, Vickers TA, Okunola HL, Bennett CF, Krainer AR. *Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice*. American journal of human genetics 2008; 82(4):834-848.

Hubbard TJ *et al*. *Ensembl 2007*. Nucleic Acids Research 2007; 35:D610-D617.

Johnson JM, Castle J, Garrett-Engle P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. *Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays*. Science 2003; 302(5653):2141-2144.

Jurica MS, Moore MJ. *Pre-mRNA Splicing: Awash in a Sea of Proteins*. Molecular Cell 2003; 12:5-14.

Karsten S, Voskoboeva E, Tishkanina S, Pettersson U, Krasnopolskaja X, Bondeson ML. *Mutational spectrum of the iduronate-2-sulfatase (IDS) gene in 36 unrelated Russian MPS II patients*. Human Genetics 1998; 103(6):732-735.

Kashima T, Manley JL. *A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy*. Nature Genetics 2003; 34:460-463.

Kendrem J. *The Encyclopedia of Molecular Biology*. Great Britain: Blackwell Science; 1994.

Kornblihtt AR. *Promoter usage and alternative splicing*. Current Opinion in cell biology 2005; 17(3):262-268.

Kráľovicová J, Vorechovský I. *Global control of aberrant splice-site activation by auxiliary splicing sequences: evidence for a gradient in exon and intron definition*. Nucleic Acids Research 2007; 35(19):6399-6413.

Krecic AM, Swanson MS. *hnRNP complexes: composition, structure, and function*. Current Opinion in Cell Biology 1999; 11(3):363-371.

Ladd AN, Cooper TA. *Finding signals that regulate alternative splicing in the post-genomic era*. Genome Biology 2002; 3(11):reviews0008.1-0008.16.

Lewis R. Part Six - Genetic Technology. Chapter 17 - *Genetic Engineering*. In *Human Genetics Concepts and Applications*. 4th edition. New York: McGrawHill; 2001. p. 325-342.

Long JC, Cáceres JF. *The SR protein family of splicing factors: master regulators of gene expression*. The biochemical journal 2009; 417:15-27.

López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. *Are splicing mutations the most frequent cause of hereditary disease?*. FEBS Letters 2005; 579(9):1900-1903.

Lualdi S, Pittis MG, Regis S, Parini R, Allegri AE, Furlan F, Bembi B, Filocamo M. *Multiple cryptic splice sites can be activated by IDS point mutations generating misspliced transcripts*. Journal of molecular medicine 2006; 84(8):692-700.

Makeyev AV, Liehaber SA. *Identification of two novel mammalian genes establishes a subfamily of KH-domain RNA-binding proteins*. Genomics 2000; 67:301–316.

Manley JL, Tacke R. *SR proteins and splicing control*. Genes & Development 1996; 10(13):1569-79.

Matlin AJ, Clark F, Smith CWJ. *Understanding alternative splicing: Towards a cellular code*. Nature Reviews. Molecular Cell Biology 2005; 6(5):386-398.

Meikle PJ, Hopwood JJ. *Lysosomal storage disorders: emerging therapeutic options require early diagnosis*. European journal of pediatrics 2003; 162(1):S34-S37.

Meikle PJ, Grasby DJ, Dean CJ, Lang DL, Bockmann M, Whittle AM, Fietz MJ, Simonsen H, Fuller M, Brooks DA, Hopwood JJ. *Newborn screening for lysosomal storage disorders*. Molecular Genetics and Metabolism 2006; 88(4):307-314.

Nilsen TW. *The spliceosome: the most complex macromolecular machine in the cell?*. BioEssays 2003; 25:1147-1149.

Pagani F, Baralle FE. *Genomic variants in exons and introns: identifying the splicing spoilers*. Nature Reviews Genetics 2004; 5:389-396.

Pagani F, Raponi M, Baralle FE. *Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution*. Proceedings of the National Academy of Sciences of the United States of America 2005; 102:6368—6372. (29 de 46)

Pajares MJ, Ezponda T, Catena R, Calvo A, Pio R, Montuenga LM. *Alternative splicing: an emerging topic in molecular and clinical oncology*. The lancet oncology 2007; 8(4):349-357.

- Pinto R, Caseiro C, Lemos M, Lopes L, Fontes A, Ribeiro H, Pinto E, Silva E, Rocha S, Marcão A, Ribeiro I, Lacerda L, Ribeiro G, Amaral O, Sá Miranda MC. *Prevalence of lysosomal storage diseases in Portugal*. European Journal of Human Genetics 2004; 12:87-92.
- Query CC. *Structural biology: Spliceosome subunit revealed*. Nature 2009; 458:418-419.
- Reed R. *Initial splice-site recognition and pairing during pre-mRNA splicing*. Current Opinion in Genetic & Development 1996; 6(2):215-220.
- Reese MG., Eeckman FH., Kulp D. e Haussler D. *Improved Splice Site Detection in Genie*. Journal of Computational Biology 1997; 4:311-23.
- Sanger F, Nicklen S, Coulson AR. *DNA sequencing with chain-terminating inhibitors*. Proceedings of the National Academy of Sciences of the United States of America 1977; 74(12):5463-5467.
- Savkur RS, Philips AV, Cooper TA. *Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy*. Nature genetics 2001; 1:40-47.
- Schellenberg MJ, Ritchie DB, MacMillan AM. *Pre-mRNA splicing: a complex picture in higher definition*. Trends in Biochemical Sciences 2008; 33(6):243-246
- Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. *Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity*. Cell 2000; 101(6):671-684.
- Singer RH, Green MR. *Compartmentalization of Eukaryotic Gene Expression: Causes and Effects*. Cell 1997; 91:291-294.
- Singh G, Cooper TA. *Minigene reporter for identification and analysis of cis elements and trans factors affecting pre-mRNA splicing*. BioTechniques 2006; 41:177-181.

Singh NN, Androphy EJ, Singh RN. *In vivo selection reveals combinatorial controls that define a critical exon in the spinal muscular atrophy genes*. RNA 2004; 10(8):1291-1305.

Singh NN, Singh RN, Androphy EJ. *Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes*. Nucleic Acids Research 2007; 35(2):371-389.

Singh NN, Shishimorova M, Cao LC, Gangwani L, Singh RN. *A short antisense oligonucleotide masking a unique intronic motif prevents skipping of a critical exon in spinal muscular atrophy*. RNA biology 2009; 6(3):341-350.

Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H. *Function of alternative splicing*. Gene 2005; 344:1-20.

Stamm S, Riethoven JJ, Le Texier V, Gopalakrishnan C, Kumanduri V, Tang Y, Barbosa-Morais NL, Thanaraj TA. *ASD: a bioinformatics resource on alternative splicing*. Nucleic Acids Research 2006; 34:D46-55

Sterner DA, Carlo T, Berget SM. *Architectural limits on split genes*. Proceedings of the National Academy of Sciences of the United States of America 1996; 93(26):15081-15085.

Tardiff DF, Lacadie SA, Rosbash M. *A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional*. Molecular Cell 2006; 24(6):917-929.

The Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff - <http://www.hgmd.cf.ac.uk/ac/index.php>.

van Deutekom JC, Janson AA, Ginjaar IB, Frankhuizen WS, Aartsma-Rus A, Bremmer-Bout M, den Dunnen JT, Koop K, van der Kooi AJ, Goemans NM, de Kimpe SJ, Ekhardt PF, Venneker EH, Platenburg GJ, Verschuuren JJ, van Ommen GJ. *Local dystrophin restoration with antisense oligonucleotide PRO051*. The New England Journal of Medicine 2007; 357(26):2677-2686.

Vega AI, Pérez-Cerdá C, Desviat LR, Matthijs G, Ugarte M, Pérez B. *Functional analysis of three splicing mutations identified in the PMM2 Gene: Toward a new therapy for congenital disorder of glycosylation type IA*. Human Mutation 2009; 30(5):795-803.

Vorechovský I. *Aberrant 3'splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization*. Nucleic Acids Research 2006; 34(16):4630-4641.

Wang GS, Cooper TA. *Splicing in disease: disruption of the splicing code and the decoding machinery*. Nature Reviews Genetics 2007; 8:749-761.

Wang Z, Burge CB. *Splicing regulation: From a parts list of regulatory elements to an integrated splicing code*. RNA 2008; 14: 802-813.

Will CL, Lührmann R. *Splicing of a rare class of introns by the U12-dependent spliceosome*. Biological chemistry 2005; 386(8):713-724.

Woolaway K, Asai K, Emili A, Cochrane A. *hnRNP E1 and E2 have distinct roles in modulating HIV-1 gene expression*. Retrovirology 2007; 4:28-46.

Xiao X, Wang Z, Jang M, Burge CB. *Coevolutionary networks of splicing cis-regulatory elements*. Proceedings of the National Academy of Sciences of the United States of America 2007; 104(47):18583-18588.

Yeo G, Burge CB. *Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals*. Journal of Computational Biology 2004; 11(2-3):377-394.

Zhang XH, Chasin LA. *Computational definition of sequence motifs governing constitutive exon splicing*. Genes & Development 2004; 18(11):1241-1250.

Zhang XH, Kangsamaksin T, Chao MS, Banerjee JK, Chasin LA. 2005. *Exon inclusion is dependent on predictable exonic splicing enhancers*. Molecular and Cellular Biology 2005; 25(16):7323-7332.

Zhu J, Mayeda A, Krainer A. *Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins*. Molecular Cell 2001; 8(6):1351-1361.